

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:11:10 ; Search time 165 Seconds
(without alignments)

560.217 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225

Sequence: 1 MATPHINAEAGDFADVILMP.....TTFNDMIKIALESVLLGDKE 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1225	100.0	239	6	AAE37217	Aae37217 Escherich
2	1224	99.9	239	6	ABR44457	AbR44457 Escherich
3	1222	99.8	239	4	AAB67584	Aab67584 Amino aci
4	1222	99.8	239	6	AAE35397	Aae35397 Escherich
5	1222	99.8	239	8	ADL27842	Adl27842 E coli de
6	1221	99.7	239	6	ABR44456	AbR44456 Escherich
7	1221	99.7	239	6	AAE35404	Aae35404 Escherich
8	1220	99.6	239	6	AAE35402	Aae35402 Escherich
9	1219	99.5	239	6	ABR44462	AbR44462 Escherich
10	1219	99.5	239	6	ABR44458	AbR44458 Escherich
11	1219	99.5	239	6	ABR44472	AbR44472 Escherich
12	1219	99.5	239	6	ABR44467	AbR44467 Escherich
13	1219	99.5	239	6	AAE35405	Aae35405 Escherich
14	1219	99.5	239	6	AAE35416	Aae35416 Escherich
15	1219	99.5	239	6	AAE35411	Aae35411 Escherich
16	1219	99.5	239	6	AAE35406	Aae35406 Escherich
17	1218	99.4	239	6	AAE37218	Aae37218 Escherich
18	1218	99.4	239	6	ABR44469	AbR44469 Escherich
19	1218	99.4	239	6	AAE35399	Aae35399 Escherich
20	1218	99.4	239	6	AAE35410	Aae35410 Escherich
21	1217	99.3	239	6	ABR44475	AbR44475 Escherich
22	1217	99.3	239	6	ABR44461	AbR44461 Escherich
23	1217	99.3	239	6	ABR44464	AbR44464 Escherich
24	1217	99.3	239	6	ABR44463	AbR44463 Escherich
25	1217	99.3	239	6	AAE35401	Aae35401 Escherich

ALIGNMENTS

RESULT 1

AAE37217
ID AAE37217 standard; protein; 239 AA.

AC AAE37217;

DT 07-AUG-2003 (first entry)

DE Escherichia coli mutant PNP enzyme (M65V) #SEQ ID 2.

KW Purine nucleoside phosphorylase; mutein; tumour; PNP; enzyme; cancer;
KW leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.

OS Escherichia coli.

OS Synthetic.

FH Key Location/Qualifiers

FT Misc-difference 65

FT /note= "this residue is Met in the wild-type sequence"

XX WO2003035012-A2.

XX 01-MAY-2003.

XX 28-OCT-2002; 2002WO-US034626.

XX 26-OCT-2001; 2001US-00035300.

XX (UABR-) UAB RES FOUND.

XX (SOOR) SOUTHERN RES INST.

XX (CORR) CORNELL RES FOUND INC.

XX Ealick SE, Parker WB, Seerist JA, Sorscher EJ;

XX WPI; 2003-421350/39.

XX N-PSDB; AAD56237.

XX Novel purified mutant purine cleaving enzyme having different biological activity than a wild-type purine cleaving enzyme, useful for treating a pathological condition characterized by abnormal cell growth.

XX Claim 9; Page 64-65; 67pp; English.

XX The invention relates to mutant purine nucleoside phosphorylase (PNP) enzymes and nucleic acid molecules encoding such enzymes. These mutant enzymes have greater biological activity than wild-type enzymes. PNP enzymes are useful for treating a pathological condition characterised by abnormal cell growth such as that occurring in cancer of the skin,

Aae35403 Escherich
Aae35412 Escherich
AbR44460 Escherich
AbR44465 Escherich
AbR44466 Escherich
AbR44471 Escherich
Aae35398 Escherich
Aae35413 Escherich
Aae35409 Escherich
Aae35415 Escherich
Aae35407 Escherich
AbR44470 Escherich
AbR44468 Escherich
AbR44473 Escherich
AbR44474 Escherich
Aae35408 Escherich
Aab67590 Amino aci
Aae35414 Escherich
Abo62285 Klebsiell

CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
 CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
 CC thyroid, testicle and uterus as well as other conditions characterised by
 CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
 CC lymphoma. Sequences of the invention are useful for treating metastatic
 CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
 CC to activate prodrugs in the treatment of cancer. They are useful for
 CC treating virally infected cells. They are also useful in gene therapy.
 CC The present sequence is Escherichia coli mutant PNP enzyme (M65V) amino
 CC acid sequence

XX
 SQ Sequence 239 AA;

Query Match 100.0%; Score 1225; DB 6; Length 239;
 Best Local Similarity 100.0%; Pred. No. 1.1e-123;
 Matches 239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMGLGTTGYGKRKISV 60
 DB 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMGLGTTGYGKRKISV 60
 QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPVHKLRDVVIGMGACTDSKVNRI 120
 DB 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPVHKLRDVVIGMGACTDSKVNRI 120
 QY 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGFVMEKYGILGVE 180
 DB 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGFVMEKYGILGVE 180
 QY 181 MEAGIYGVAEFGAKALITCTVSDHIRTHTQTTAAERQTTFNMIKIALESVLLGDKE 239
 DB 181 MEAGIYGVAEFGAKALITCTVSDHIRTHTQTTAAERQTTFNMIKIALESVLLGDKE 239

RESULT 2

ABR44457
 ID ABR44457 standard; protein; 239 AA.

XX AC ABR44457;

XX 07-AUG-2003 (first entry)

XX Escherichia coli mutant PNP enzyme M65I.

XX Purine nucleoside phosphorylase; murein; tumour; PNP; enzyme; cancer;
 KW leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.

XX Escherichia coli.

OS Synthetic.

XX Key Location/Qualifiers

PH Misc-difference 65
 FT /note= "wild-type Met replaced with Ile"

XX WO2003035012-A2.

XX 01-MAY-2003.

XX 28-OCT-2002; 2002WO-US034626.

XX 26-OCT-2001; 2001US-00035300.

XX (UABR-) UAB RES FOUND.

PA (SOUR) SOUTHERN RES INST.

PA (CORR) CORNELL RES FOUND INC.

XX Balick SE, Parker WB, Seerist JA, Sorscher EJ;

XX WPI; 2003-421350/39.

XX Novel purified mutant purine cleaving enzyme having different biological
 PT activity than a wild-type purine cleaving enzyme, useful for treating a
 PT pathological condition characterized by abnormal cell growth.

XX Claim 13; Page; 67pp; English.
 PS
 XX The invention relates to mutant purine nucleoside phosphorylase (PNP)
 CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
 CC enzymes have greater biological activity than wild-type enzymes. PNP
 CC enzymes are useful for treating a pathological condition characterised by
 CC abnormal cell growth such as that occurring in cancer of the skin,
 CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
 CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
 CC thyroid, testicle and uterus as well as other conditions characterised by
 CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
 CC lymphoma. Sequences of the invention are useful for treating metastatic
 CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
 CC to activate prodrugs in the treatment of cancer. They are useful for
 CC treating virally infected cells. They are also useful in gene therapy.
 CC The present sequence is Escherichia coli mutant PNP enzyme (M65I) amino
 CC acid sequence. NOTE: This sequence is given in the specification, but may
 CC be created by making the wild-type sequence from the mutant sequence
 CC given in SEQ ID 2 (AAB37217) and then creating the desired mutation

XX Sequence 239 AA;

Query Match 99.9%; Score 1224; DB 6; Length 239;
 Best Local Similarity 99.6%; Pred. No. 1.5e-123;
 Matches 239; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMGLGTTGYGKRKISV 60
 DB 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMGLGTTGYGKRKISV 60
 QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPVHKLRDVVIGMGACTDSKVNRI 120
 DB 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPVHKLRDVVIGMGACTDSKVNRI 120
 QY 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGFVMEKYGILGVE 180
 DB 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGFVMEKYGILGVE 180
 QY 181 MEAGIYGVAEFGAKALITCTVSDHIRTHTQTTAAERQTTFNMIKIALESVLLGDKE 239
 DB 181 MEAGIYGVAEFGAKALITCTVSDHIRTHTQTTAAERQTTFNMIKIALESVLLGDKE 239

RESULT 3

AAB57584
 ID AAB57584 standard; protein; 239 AA.

XX AC AAB57584;

XX 29-MAY-2001 (first entry)

XX Amino acid sequence of a purine nucleoside phosphorylase.

XX Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;
 KW purine nucleoside phosphorylase; phosphopentose mutase;
 KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;
 KW deoxyribokinase; nucleoside 2-deoxyriboseyltransferase.

XX Escherichia coli.

XX WO200114566-A2.

XX 01-MAR-2001.

XX 18-AUG-2000; 2000WO-EP008088.

XX 20-AUG-1999; 99EP-00116425.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

PA (INSP) INST PASTEUR.

PA (PHAR-) PHARMA-WALDHOF GMBH & CO KG.

XX

PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;
PI Marliere P, Pochet S;

DR WPI; 2001-235026/24.

DR N-PSDB; AAP55440.

XX In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting
PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside
PT and an inorganic phosphate.

XX Disclosure; Page 45; 73pp; English.

XX The present sequence represents a purine nucleoside phosphorylase enzyme.
CC This enzyme is involved in the biosynthesis of deoxyribonucleosides, and
CC is used in the method of the invention. The specification describes a
CC method for the in vitro enzymatic synthesis of deoxyribonucleosides. The
CC method comprises reacting deoxyribose 1-phosphate and a nucleobase to
CC form a deoxyribonucleoside and an inorganic phosphate. Enzymes which may
CC be used in the method of the invention include thymidine phosphorylase,
CC purine nucleoside phosphorylase, phosphopentose mutase, phosphopentose
CC aldolase, fructose 1,6-diphosphate aldolase, deoxyribokinase, and
CC nucleoside 2-deoxyribosyltransferase

XX Sequence 239 AA;

Query Match 99.8%; Score 1222; DB 4; Length 239;

Best Local Similarity 99.6%; Pred. No. 2.4e-123;

Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEAGDPADVVLMPGDLPRAKYIAETFLDAREVNNVRGMLGFTGTGKRSV 60

Db 1 MATPHINAEAGDPADVVLMPGDLPRAKYIAETFLDAREVNNVRGMLGFTGTGKRSV 60

QY 61 MGHGVIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGIGACTDSKVNIR 120

Db 61 MGHGVIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGIGACTDSKVNIR 120

QY 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVKYILGVE 180

Db 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVKYILGVE 180

QY 181 MEAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239

Db 181 MEAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239

RESULT 4

AAE35397
ID AAE35397 standard; protein; 239 AA.

AC AAE35397;

DT 17-JUN-2003 (first entry)

XX Escherichia coli cytosine nucleoside phosphorylase.

XX Cytosine nucleoside phosphorylase; cytosine nucleoside compound;
KW pharmaceutical; enzyme.

XX Escherichia coli.

OS Escherichia coli.

XX EP1254959-A2.

PN 06-NOV-2002.

XX 01-MAY-2002; 2002EP-00253075.

XX 01-MAY-2001; 2001JP-00134352.

XX (MITA) MITSUI CHEM INC.

XX Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;

PI Ishibaishi H, Nagahara K, Fukui Y;

XX WPI; 2003-158787/16.

DR N-PSDB; AAD54112.

XX Producing cytosine nucleoside compound by reacting sugar phosphate and
PT cytosine or cytosine derivative in the presence of cytosine nucleoside
PT phosphorylase.

PS Claim 11; Page 18; 32pp; English.

XX The invention relates to a method for producing cytosine nucleoside
CC compound by reacting sugar phosphate and cytosine or cytosine derivative
CC in the presence of cytosine nucleoside phosphorylase. The method is used
CC to produce cytosine nucleoside compounds which are intermediates for
CC pharmaceuticals. The present sequence is Escherichia coli cytosine
CC nucleoside phosphorylase

XX Sequence 239 AA;

Query Match 99.8%; Score 1222; DB 6; Length 239;

Best Local Similarity 99.6%; Pred. No. 2.4e-123;

Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEAGDPADVVLMPGDLPRAKYIAETFLDAREVNNVRGMLGFTGTGKRSV 60

Db 1 MATPHINAEAGDPADVVLMPGDLPRAKYIAETFLDAREVNNVRGMLGFTGTGKRSV 60

QY 61 MGHGVIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGIGACTDSKVNIR 120

Db 61 MGHGVIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVGIGACTDSKVNIR 120

QY 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVKYILGVE 180

Db 121 FKDHFAAIAADFDMVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVKYILGVE 180

QY 181 MEAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239

Db 181 MEAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239

RESULT 5

ADL27842
ID ADL27842 standard; protein; 239 AA.

XX ADL27842;

DT 20-MAY-2004 (first entry)

XX E coli deod.

XX selection method; yield; pathogen resistance; nutritional quality;
KW stress resistance; plant; transgenic.

XX Escherichia coli.

XX WO2004013333-A2.

XX 12-FEB-2004.

PF 18-JUL-2003; 2003WO-EP007877.

XX 26-JUL-2002; 2002DE-01034287.

XX (BADI) BASF PLANT SCI GMBH.

XX Kock M, Frank M, Badur R;

XX WPI; 2004-157134/15.

XX N-PSDB; ADL27841.

XX Method for transforming plant cells, useful for preparing transgenic
PT plants for e.g. food or production of chemicals, with selection based on
PT suppressing toxic effects of a marker protein.

XX PS Claim 6; Page 140-141; 201pp; German.

XX CC The present invention relates to a method for preparing transformed plant

CC cells or organisms, which comprises first transfecting a population of

CC plant cells, already containing at least one marker protein that can

CC exert a direct or indirect toxic effect, with at least one nucleic acid

CC sequence together with at least one compound able to reduce expression,

CC amount, activity and/or function of the marker protein. Transfected cells

CC that contain the nucleic acid in their genome and, because of the

CC activity of the compound have a growth advantage relative to non-

CC transformed cells are selected, under conditions where the marker protein

CC exerts its toxic effect on non-transformed cells. The method is

CC especially used to produce transgenic plants, especially those having an

CC advantageous phenotype, e.g. better nutritional quality for humans or

CC animals, production of selected chemicals or pharmaceuticals, increased

CC resistance to pathogens or environmental stress or higher yields. The

CC present sequence is a polypeptide shown in the exemplification of the

CC invention.

XX CC

XX SQ Sequence 239 AA;

Query Match 99.8%; Score 1222; DB 8; Length 239;

Best Local Similarity 99.6%; Pred. No. 2.4e-123;

Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVLMPCDPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60

Db 1 MATPHINAEMGDFADVLMPCDPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIWVGACTDSKVNRI 120

Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIWVGACTDSKVNRI 120

Qy 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180

Db 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180

Qy 181 MEAGIYGVAAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239

Db 181 MEAGIYGVAAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239

RESULT 6

ABR44456

ID ABR44456 standard; protein; 239 AA.

XX AC ABR44456;

XX XX

XX DT 07-AUG-2003 (first entry)

XX DE Escherichia coli mutant PNP enzyme M65A.

XX CC Purine nucleoside phosphorylase; mutein; tumour; PNP; enzyme; cancer;

KW leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.

XX CC

OS Escherichia coli.

OS Synthetic.

XX CC

XX FH Key Location/Qualifiers

FT Misc-difference 65

FT /note= "wild-type Met replaced with Ala"

XX PN WO2003035012-A2.

XX XX

XX PD 01-MAY-2003.

XX PF 28-OCT-2002; 2002WO-US034626.

XX PR 26-OCT-2001; 2001US-00035300.

XX PA (UABR-) UAB RES FOUND.

PA (SOUR) SOUTHERN RES INST.

PA (CORR) CORNELL RES FOUND INC.

XX PI Ealick SE, Parker WB, Seerist JA, Sorscher EJ;

XX DR WPI; 2003-421350/39.

XX PT Novel purified mutant purine cleaving enzyme having different biological

PT activity than a wild-type purine cleaving enzyme, useful for treating a

PT pathological condition characterized by abnormal cell growth.

XX PS Claim 13; Page; 67pp; English.

XX CC The invention relates to mutant purine nucleoside phosphorylase (PNP)

CC enzymes and nucleic acid molecules encoding such enzymes. These mutant

CC enzymes have greater biological activity than wild-type enzymes. PNP

CC enzymes are useful for treating a pathological condition characterised by

CC abnormal cell growth such as that occurring in cancer of the skin,

CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,

CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,

CC thyroid, testicle and uterus as well as other conditions characterised by

CC abnormal cell growth such as myeloid leukaemia, glioblastoma and

CC lymphoma. Sequences of the invention are useful for treating metastatic

CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or

CC to activate prodrugs in the treatment of cancer. They are useful for

CC treating virally infected cells. They are also useful in gene therapy.

CC The present sequence is Escherichia coli mutant PNP enzyme (M65A) amino

CC acid sequence. NOTE: This sequence is given in the specification, but may

CC be created by making the wild-type sequence from the mutant sequence

CC given in SEQ ID 2 (AAE37217) and then creating the desired mutation

XX CC

XX SQ Sequence 239 AA;

Query Match 99.7%; Score 1221; DB 6; Length 239;

Best Local Similarity 99.6%; Pred. No. 3.1e-123;

Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVLMPCDPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60

Db 1 MATPHINAEMGDFADVLMPCDPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIWVGACTDSKVNRI 120

Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIWVGACTDSKVNRI 120

Qy 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180

Db 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180

Qy 181 MEAGIYGVAAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239

Db 181 MEAGIYGVAAEFGAKALITCTVSDHIRTHTHEQTAAERQTTFNDMIKIALESVLLGDKE 239

RESULT 7

AAE35404

ID AAE35404 standard; protein; 239 AA.

XX AC AAE35404;

XX XX

XX DT 17-JUN-2003 (first entry)

XX DE Escherichia coli cytosine nucleoside phosphorylase mutant, V1041.

XX CC Cytosine nucleoside phosphorylase; cytosine nucleoside compound;

KW pharmaceutical; enzyme; mutant; mutein.

XX CC

OS Escherichia coli.

OS Synthetic.

XX CC

XX FH Key Location/Qualifiers

FT Misc-difference 104

FT /note= "wild-type Val is substituted with Ile"

XX CC


```
PN EP1254959-A2.
XX
XX PD 06-NOV-2002.
XX
XX PF 01-MAY-2002; 2002EP-00253075.
XX
XX PR 01-MAY-2001; 2001JP-00134352.
XX
XX PA (MITA ) MITSUI CHEM INC.
XX
XX PI Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;
XX Ishibashi H, Nagahara K, Fukui Y;
XX WPI; 2003-158787/16.
XX
XX PT Producing cytosine nucleoside compound by reacting sugar phosphate and
XX cytosine or cytosine derivative in the presence of cytosine nucleoside
XX phosphorylase.
XX
XX PS Example 13; Page; 32pp; English.
XX
XX CC The invention relates to a method for producing cytosine nucleoside
XX compound by reacting sugar phosphate and cytosine or cytosine derivative
XX in the presence of cytosine nucleoside phosphorylase. The method is used
XX to produce cytosine nucleoside compounds which are intermediates for
XX pharmaceuticals. The present sequence is Escherichia coli cytosine
XX nucleoside phosphorylase. Note: This sequence is not shown in the
XX specification, however it is constructed based on E. coli cytosine
XX nucleoside phosphorylase (AAE35397) sequence shown in column 18
XX
XX SQ Sequence 239 AA;
Query Match 99.7%; Score 1221; DB 6; Length 239;
Best Local Similarity 99.2%; Pred. No. 3.1e-123;
Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATPHINAEAGDFADVLMVPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
DB 1 MATPHINAEAGDFADVLMVPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGIGACTDSKVNIR 120
DB 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGIGACTDSKVNIR 120
QY 121 FKDHDFAAIADFDVNRNAVDAKALGIDARVGNLFSADLFYSPDGEMFVMEKYGILGVE 180
DB 121 FKDHDFAAIADFDVNRNAVDAKALGIDARVGNLFSADLFYSPDGEMFVMEKYGILGVE 180
QY 181 MEAAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLGDKE 239
DB 181 MEAAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLGDKE 239
RESULT 8
AAE35402
ID AAE35402 standard; protein; 239 AA.
XX
XX AC AAE35402;
XX
XX DT 17-JUN-2003 (first entry)
XX
XX DE Escherichia coli cytosine nucleoside phosphorylase mutant, I67L.
XX
XX KW Cytosine nucleoside phosphorylase; cytosine nucleoside compound;
XX pharmaceutical; enzyme; mutant; muten.
XX
XX OS Escherichia coli.
XX Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 67
XX FT /note= "wild-type Ile is substituted with Leu"
```

```
PN EP1254959-A2.
XX
XX PD 06-NOV-2002.
XX
XX PF 01-MAY-2002; 2002EP-00253075.
XX
XX PR 01-MAY-2001; 2001JP-00134352.
XX
XX PA (MITA ) MITSUI CHEM INC.
XX
XX PI Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;
XX Ishibashi H, Nagahara K, Fukui Y;
XX WPI; 2003-158787/16.
XX
XX PT Producing cytosine nucleoside compound by reacting sugar phosphate and
XX cytosine or cytosine derivative in the presence of cytosine nucleoside
XX phosphorylase.
XX
XX PS Example 13; Page; 32pp; English.
XX
XX CC The invention relates to a method for producing cytosine nucleoside
XX compound by reacting sugar phosphate and cytosine or cytosine derivative
XX in the presence of cytosine nucleoside phosphorylase. The method is used
XX to produce cytosine nucleoside compounds which are intermediates for
XX pharmaceuticals. The present sequence is Escherichia coli cytosine
XX nucleoside phosphorylase. Note: This sequence is not shown in the
XX specification, however it is constructed based on E. coli cytosine
XX nucleoside phosphorylase (AAE35397) sequence shown in column 18
XX
XX SQ Sequence 239 AA;
Query Match 99.6%; Score 1220; DB 6; Length 239;
Best Local Similarity 99.2%; Pred. No. 3.9e-123;
Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATPHINAEAGDFADVLMVPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
DB 1 MATPHINAEAGDFADVLMVPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYKGRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGIGACTDSKVNIR 120
DB 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVGIGACTDSKVNIR 120
QY 121 FKDHDFAAIADFDVNRNAVDAKALGIDARVGNLFSADLFYSPDGEMFVMEKYGILGVE 180
DB 121 FKDHDFAAIADFDVNRNAVDAKALGIDARVGNLFSADLFYSPDGEMFVMEKYGILGVE 180
QY 181 MEAAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLGDKE 239
DB 181 MEAAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNDMIKIALESVLLGDKE 239
RESULT 9
ABR44462
ID ABR44462 standard; protein; 239 AA.
XX
XX AC ABR44462;
XX
XX DT 07-AUG-2003 (first entry)
XX
XX DE Escherichia coli mutant PNP enzyme E180D.
XX
XX KW Purine nucleoside phosphorylase; muten; tumour; PNP; enzyme; cancer;
XX leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.
XX
XX OS Escherichia coli.
XX Synthetic.
XX
XX FH Key Location/Qualifiers
XX FT Misc-difference 180
XX FT /note= "wild-type Glu replaced with Asp"
```

PN W02003035012-A2.
PD 01-MAY-2003.
XX 28-OCT-2002; 2002WO-US034626.
XX 26-OCT-2001; 2001US-00035300.
XX (UABR-) UAB RES FOUND.
PA (SOUR) SOUTHERN RES INST.
FA (CORR) CORNELL RES FOUND INC.
XX
PI Ealick SE, Parker WB, Seerist JA, Sorscher EJ;
XX WPI; 2003-421350/39.
DR
XX Novel purified mutant purine cleaving enzyme having different biological
PT activity than a wild-type purine cleaving enzyme, useful for treating a
PT pathological condition characterized by abnormal cell growth.
XX
PS Claim 13; Page; 67pp; English.
XX
CC The invention relates to mutant purine nucleoside phosphorylase (PNP)
CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
CC enzymes have greater biological activity than wild-type enzymes. PNP
CC enzymes are useful for treating a pathological condition characterised by
CC abnormal cell growth such as that occurring in cancer of the skin,
CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
CC thyroid, testicle and uterus as well as other conditions characterised by
CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
CC lymphoma. Sequences of the invention are useful for treating metastatic
CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
CC to activate prodrugs in the treatment of cancer. They are useful for
CC treating virally infected cells. They are also useful in gene therapy.
CC The present sequence is Escherichia coli mutant PNP enzyme (E180D) amino
CC acid sequence. NOTE: This sequence is given in the specification, but may
CC be created by making the wild-type sequence from the mutant sequence
CC given in SEQ ID 2 (AAE37217) and then creating the desired mutation
XX
SQ Sequence 239 AA;
Query Match 99.5%; Score 1219; DB 6; Length 239;
Best Local Similarity 99.2%; Pred. No. 5.1e-123;
Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MATPHINAEMGDFADVLLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTCTYGRKISV 60
Db 1 MATPHINAEMGDFADVLLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTCTYGRKISV 60
Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVIWVGACTDSKVNRI 120
Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVIWVGACTDSKVNRI 120
Qy 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYGILGVE 180
Db 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYGILGVD 180
Qy 181 MEAGIYGVAAEFCAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDK 239
Db 181 MEAGIYGVAAEFCAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDK 239
RESULT 10
ABR44458 standard; protein; 239 AA.
XX ABR44458;
AC ABR44458;
XX
DT 07-AUG-2003 (first entry)
XX Escherichia coli mutant PNP enzyme M65Q.
DE
XX

KW Purine nucleoside phosphorylase; mutein; tumour; PNP; enzyme; cancer;
KW leukaemia; glioblastoma; lymphoma; carcinoma; gene therapy.
OS Escherichia coli.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 65 /note= "wild-type Met replaced with Gln"
FT
XX W02003035012-A2.
XX 01-MAY-2003.
XX 28-OCT-2002; 2002WO-US034626.
XX 26-OCT-2001; 2001US-00035300.
XX (UABR-) UAB RES FOUND.
PA (SOUR) SOUTHERN RES INST.
PA (CORR) CORNELL RES FOUND INC.
XX
XX Ealick SE, Parker WB, Seerist JA, Sorscher EJ;
XX WPI; 2003-421350/39.
DR
XX Novel purified mutant purine cleaving enzyme having different biological
PT activity than a wild-type purine cleaving enzyme, useful for treating a
PT pathological condition characterized by abnormal cell growth.
XX
PS Claim 13; Page; 67pp; English.
XX
CC The invention relates to mutant purine nucleoside phosphorylase (PNP)
CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
CC enzymes have greater biological activity than wild-type enzymes. PNP
CC enzymes are useful for treating a pathological condition characterised by
CC abnormal cell growth such as that occurring in cancer of the skin,
CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
CC thyroid, testicle and uterus as well as other conditions characterised by
CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
CC lymphoma. Sequences of the invention are useful for treating metastatic
CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
CC to activate prodrugs in the treatment of cancer. They are useful for
CC treating virally infected cells. They are also useful in gene therapy.
CC The present sequence is Escherichia coli mutant PNP enzyme (M65Q) amino
CC acid sequence. NOTE: This sequence is given in the specification, but may
CC be created by making the wild-type sequence from the mutant sequence
CC given in SEQ ID 2 (AAE37217) and then creating the desired mutation
XX
SQ Sequence 239 AA;
Query Match 99.5%; Score 1219; DB 6; Length 239;
Best Local Similarity 99.6%; Pred. No. 5.1e-123;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MATPHINAEMGDFADVLLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTCTYGRKISV 60
Db 1 MATPHINAEMGDFADVLLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTCTYGRKISV 60
Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVIWVGACTDSKVNRI 120
Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVIWVGACTDSKVNRI 120
Qy 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYGILGVE 180
Db 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYGILGVE 180
Qy 181 MEAGIYGVAAEFCAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDK 239
Db 181 MEAGIYGVAAEFCAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDK 239

RESULT 11

ABR44472
ID ABR44472 standard; protein; 239 AA.

AC ABR44472;

XX 07-AUG-2003 (first entry)

DE Escherichia coli mutant PNP enzyme E182Q.

KW Purine nucleoside phosphorylase; mutain; tumour; PNP; enzyme; cancer;
KW leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.

XX Escherichia coli.

OS Synthetic.

PH Key Location/Qualifiers

FT Misc-difference 182
FT /note= "wild-type Glu replaced with Gln"

XX WO2003035012-A2.

PD 01-MAY-2003.

PF 28-OCT-2002; 2002WO-US034626.

PR 26-OCT-2001; 2001US-00035300.

XX (UABR-) UAB RES FOUND.

PA (SOUR) SOUTHERN RES INST.

PA (CORR) CORNELL RES FOUND INC.

PI Ealick SE, Parker WB, Seerist JA, Sorscher EJ;

XX WPI; 2003-421350/39.

XX Novel purified mutant purine cleaving enzyme having different biological
PT activity than a wild-type purine cleaving enzyme, useful for treating a
PT pathological condition characterized by abnormal cell growth.

PS Claim 13; Page; 67pp; English.

XX The invention relates to mutant purine nucleoside phosphorylase (PNP)
CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
CC enzymes have greater biological activity than wild-type enzymes. PNP
CC enzymes are useful for treating a pathological condition characterised by
CC abnormal cell growth such as that occurring in cancer of the skin,
CC breast, bone, colon, head or neck, kidney, lung, liver, ovary,
CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
CC thyroid, testicle and uterus as well as other conditions characterised by
CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
CC lymphoma. Sequences of the invention are useful for treating metastatic
CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
CC to activate prodrugs in the treatment of cancer. They are useful for
CC treating virally infected cells. They are also useful in gene therapy.
CC The present sequence is Escherichia coli mutant PNP enzyme (E182Q) amino
CC acid sequence. NOTE: This sequence is given in the specification, but may
CC be created by making the wild-type sequence from the mutant sequence
CC given in SEQ ID 2 (AAE37217) and then creating the desired mutation

XX Sequence 239 AA;

Query Match 99.5%; Score 1219; DB 6; Length 239;

Best Local Similarity 99.2%; Pred. No. 5, 1e-123;

Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATPHINAEAGDFADVLMVMPGDLRAKYAETFFLEDAEVNVRVGLGFTGYGRKISV 60

Db 1 MATPHINAEAGDFADVLMVMPGDLRAKYAETFFLEDAEVNVRVGLGFTGYGRKISV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVIHGACTDSKVNIR 120

Db 61 MGHGMGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVIHGACTDSKVNIR 120

Qy 121 FKDHDFAAIADFDVVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVE 180
Db 121 FKDHDFAAIADFDVVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVE 180
Qy 181 MEAAGIYGVAAEFAGAKALTICTVSDHIRTHTQTTAAERQTTTFNNDMIKIALESVLLGDKE 239
Db 181 MQAAGIYGVAAEFAGAKALTICTVSDHIRTHTQTTAAERQTTTFNNDMIKIALESVLLGDKE 239

RESULT 12

ABR44467
ID ABR44467 standard; protein; 239 AA.

XX ABR44467;

XX 07-AUG-2003 (first entry)

XX Escherichia coli mutant PNP enzyme M181L.

XX Purine nucleoside phosphorylase; mutain; tumour; PNP; enzyme; cancer;
KW leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy.

XX Escherichia coli.

OS Synthetic.

PH Key Location/Qualifiers

FT Misc-difference 181

FT /note= "wild-type Met replaced with Leu"

XX WO2003035012-A2.

XX 01-MAY-2003.

XX 28-OCT-2002; 2002WO-US034626.

XX 26-OCT-2001; 2001US-00035300.

XX (UABR-) UAB RES FOUND.

PA (SOUR) SOUTHERN RES INST.

PA (CORR) CORNELL RES FOUND INC.

PI Ealick SE, Parker WB, Seerist JA, Sorscher EJ;

XX WPI; 2003-421350/39.

XX Novel purified mutant purine cleaving enzyme having different biological
PT activity than a wild-type purine cleaving enzyme, useful for treating a
PT pathological condition characterized by abnormal cell growth.

PS Claim 13; Page; 67pp; English.

XX The invention relates to mutant purine nucleoside phosphorylase (PNP)
CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
CC enzymes have greater biological activity than wild-type enzymes. PNP
CC enzymes are useful for treating a pathological condition characterised by
CC abnormal cell growth such as that occurring in cancer of the skin,
CC breast, bone, colon, head or neck, kidney, lung, liver, ovary,
CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
CC thyroid, testicle and uterus as well as other conditions characterised by
CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
CC lymphoma. Sequences of the invention are useful for treating metastatic
CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
CC to activate prodrugs in the treatment of cancer. They are useful for
CC treating virally infected cells. They are also useful in gene therapy.
CC The present sequence is Escherichia coli mutant PNP enzyme (M181L) amino
CC acid sequence. NOTE: This sequence is given in the specification, but may
CC be created by making the wild-type sequence from the mutant sequence
CC given in SEQ ID 2 (AAE37217) and then creating the desired mutation

XX Sequence 239 AA;

Query Match 99.5%; Score 1219; DB 6; Length 239;

	Best Local Similarity	99.2%;	Pred. No. 5.1e-123;	Matches	237;	Conservative	2;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	MATHPHNAEMGDFADVVLMPGDP	LR	AKYIAETFL	EDAREVNNV	RGMLGFTCTY	YGRKISV	60					
Db	1	MATHPHNAEMGDFADVVLMPGDP	LR	AKYIAETFL	EDAREVNNV	RGMLGFTCTY	YGRKISV	60					
Qy	61	MGHGVGIPSCSIYTKELITD	FGVKKI	IRVSGCGAVL	PHVKLRD	VVVGWACTD	TSKVNRI	120					
Db	61	MGHGVGIPSCSIYTKELITD	FGVKKI	IRVSGCGAVL	PHVKLRD	VVVGWACTD	TSKVNRI	120					
Qy	121	FKDHDFAAIADPDMVRNVA	DA	AKALGIDTAR	VGNLFSADL	FSPDGE	MPDMVEKYGILGVE	180					
Db	121	FKDHDFAAIADPDMVRNVA	DA	AKALGIDTAR	VGNLFSADL	FSPDGE	MPDMVEKYGILGVE	180					
Qy	181	MEAAAGIYGVAAEFCAKALTI	CTVSDH	IRTHEQT	TAAERQTT	FNDM	IKIALESVLLGDK	239					
Db	181	LEAAGIYGVAAEFCAKALTI	CTVSDH	IRTHEQT	TAAERQTT	FNDM	IKIALESVLLGDK	239					

RESULT 13	
AAE35405	
ID	AAE35405 standard; protein; 239 AA.
XX	
AC	AAE35405;
XX	
DT	17-JUN-2003 (first entry)
XX	
DE	Escherichia coli cytosine nucleoside phosphorylase mutant, V135M.
XX	
KW	Cytosine nucleoside phosphorylase; cytosine nucleoside compound;
KW	pharmaceutical; enzyme; mutant; mutein.
XX	
OS	Escherichia coli.
OS	Synthetic.
XX	
PH	Key Location/Qualifiers
FT	Misc-difference 135
FT	/note= "Wild-type Val is substituted with Met"
XX	
FN	EP1254959-A2.
XX	
PD	06-NOV-2002.
XX	
PF	01-MAY-2002; 2002EP-00253075.
XX	
PR	01-MAY-2001; 2001JP-00134352.
XX	
PA	(MITA) MITSUI CHEM INC.
XX	
PI	Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;
PI	Ishibashi H, Nagahara K, Fukui Y;
XX	
DR	WPI; 2003-158787/16.
XX	
PT	Producing cytosine nucleoside compound by reacting sugar phosphate and
PT	cytosine or cytosine derivative in the presence of cytosine nucleoside
PT	phosphorylase.
XX	
PS	Example 13; Page; 32pp; English.
XX	
CC	The invention relates to a method for producing cytosine nucleoside
CC	compound by reacting sugar phosphate and cytosine or cytosine derivative
CC	in the presence of cytosine nucleoside phosphorylase. The method is used
CC	to produce cytosine nucleoside compounds which are intermediates for
CC	pharmaceuticals. The present sequence is Escherichia coli cytosine
CC	nucleoside phosphorylase. Note: This sequence is not shown in the
CC	specification, however it is constructed based on E. coli cytosine
CC	nucleoside phosphorylase (AAE35397) sequence shown in column 18
XX	
SQ	Sequence 239 AA;

Query Match 99.5%; Score 1219; DB 6; Length 239;

Query Match 99.5%; Score 1219; DB 6; Length 239;

Qy	Best Local Similarity	99.2%;	Pred. No. 5.1e-123;	Matches	237;	Conservative	2;	Mismatches	0;	Indels	0;	Gaps	0
Qy	1	MATPHINAEMGD	PADVILMPGD	PLRAKYIAETFFLEDAREVNNVR	GMGLFTGT	YTKRKISV	60						
Db	1	MATPHINAEMGD	PADVILMPGD	PLRAKYIAETFFLEDAREVNNVR	GMGLFTGT	YTKRKISV	60						
Qy	61	MGHGVGIPSCSI	YTKELITD	FGVKKIIRVSGCGAVLPHVKLRD	VVIGMCACTD	SKVNRIR	120						
Db	61	MGHGVGIPSCSI	YTKELITD	FGVKKIIRVSGCGAVLPHVKLRD	VVIGMCACTD	SKVNRIR	120						
Qy	121	FKDHFPAAIAD	PMVRNVA	DAKALIGIDARVGNLF	SAADLYSPDCEM	EDVMEKYILGVE	180						
Db	121	FKDHFPAAIAD	PMVRNVA	DAKALIGIDARVGNLF	SAADLYSPDCEM	EDVMEKYILGVE	180						
Qy	181	MEAGIYGVAAEF	CAKALITICTVSDH	IRTHEOTTAAERQTTFN	DMIKIALSVLLGDKE	239							
Db	181	MEAGIYGVAAEF	CAKALITICTVSDH	IRTHEOTTAAERQTTFN	DMIKIALSVLLGDKE	239							

RESULT 15	
AAE35411	
ID	AAE35411 standard; protein; 239 AA.
XX	
XX	
AC	AAE35411;
XX	
DT	17-JUN-2003 (first entry)
XX	
DE	Escherichia coli cytosine nucleoside phosphorylase mutant, A183S.
XX	
KW	Cytosine nucleoside phosphorylase; cytosine nucleoside compound;
XX	pharmaceutical; enzyme; mutant; mutein.
XX	
OS	Escherichia coli.
OS	Synthetic.
XX	
PH	Key Location/Qualifiers
FT	Misc-difference 183
FT	/note= "Wild-type Ala is substituted with Ser"
XX	
XX	
PN	EP1254959-A2.
XX	
PD	06-NOV-2002.
XX	
PF	01-MAY-2002; 2002EP-00253075.
XX	
PR	01-MAY-2001; 2001JP-00134352.
XX	
PA	(MITA) MITSUI CHEM INC.
PI	Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;
PI	Ishibashi H, Nagahara K, Fukui Y;
XX	
DR	WPI; 2003-158787/16.
XX	
PT	Producing cytosine nucleoside compound by reacting sugar phosphate and
PT	cytosine or cytosine derivative in the presence of cytosine nucleoside
PT	phosphorylase.
XX	
PS	Example 13; Page; 32pp; English.
XX	
CC	The invention relates to a method for producing cytosine nucleoside
CC	compound by reacting sugar phosphate and cytosine or cytosine derivative
CC	in the presence of cytosine nucleoside phosphorylase. The method is used
CC	to produce cytosine nucleoside compounds which are intermediates for
CC	pharmaceuticals. The present sequence is Escherichia coli cytosine
CC	nucleoside phosphorylase. Note: This sequence is not shown in the
CC	specification, however it is constructed based on E. coli cytosine
CC	nucleoside phosphorylase (AAE35397) sequence shown in column 18
XX	
SQ	Sequence 239 AA;

Query Match
99.5%; Score 1219; DB 6; Length 239;

	Best Local Similarity	99.2%; Pred. No. 5.1e-123;	Mismatches	0;	Indels	0;	Gaps	0
	Matches	23;	Conservative	2;				
Qy	1	MATPHINAEMGDFADVVLMPGDPURAKYIAFTFLIEDAREVNNVRGMLGFTGTGYKGRKISV	60					
Db	1	MATPHINAEMGDFADVVLMPGDPURAKYIAFTFLIEDAREVNNVRGMLGFTGTGYKGRKISV	60					
Qy	61	MHGVGIGIPSCSIYTKELITDGFVKIIIRVSGCGAVLPHVKLRDVVVIGMGACTDTSKVNRI	120					
Db	61	MHGGMGIPSCSIYTKELITDGFVKIIIRVSGCGAVLPHVKLRDVVVIGMGACTDTSKVNRI	120					
Qy	121	FKDHDFAAIADFDMVRNVAADAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYILGVGE	180					
Db	121	FKDHDFAAIADFDMVRNVAADAAKALGIDARVGNLFSADLFYSPDGEFDMVEKYILGVGE	180					
Qy	181	MEAGIYGVAAEFCAKALTICTVSDHIRTHTBQTTAAERQTTTFNDMIKIALESVLLGDKE	239					
Db	181	MEAGIYGVAAEFCAKALTICTVSDHIRTHTBQTTAAERQTTTFNDMIKIALESVLLGDKE	239					

Search completed: July 18, 2005, 21:36:53
Job time : 166 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 22, 2005, 15:10:32 ; Search time 5506 Seconds
(without alignments)
6336.316 Million cell updates/sec

Title: US-10-035-300A-1

Perfect score: 720
Sequence: 1 atggctaccacacattaa.....tgctggggcgataaagagtaa 720

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.on.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	716.8	99.6	720	6	BD187724	BD187724 A method
2	716.8	99.6	720	6	CQ774461	CQ774461 Sequence
3	716.8	99.6	720	6	AX087917	AX087917 Sequence
4	716.8	99.6	720	6	AX590437	AX590437 Sequence
5	716.8	99.6	1249	1	ECOPNP	M60917 E.coli puri
6	716.8	99.6	3031	6	BD261823	BD261823 Recombina
7	716.8	99.6	3031	6	AX027820	AX027820 Recombina
8	716.8	99.6	3128	6	BD261824	BD261824 Recombina
9	716.8	99.6	3128	6	AX027821	AX027821 Recombina
10	716.8	99.6	3383	6	BD261814	BD261814 Recombina
11	716.8	99.6	3383	6	AX027811	AX027811 Sequence
12	716.8	99.6	3934	6	BD261825	BD261825 Recombina
13	716.8	99.6	3934	6	AX027822	AX027822 Sequence
14	716.8	99.6	4189	6	BD261816	BD261816 Recombina
15	716.8	99.6	4189	6	AX027813	AX027813 Sequence
16	716.8	99.6	5013	6	AR264513	AR264513 Sequence
17	716.8	99.6	5241	6	BD261818	BD261818 Recombina
18	716.8	99.6	5241	6	AX027815	AX027815 Sequence
19	716.8	99.6	5495	6	BD261815	BD261815 Recombina

20	716.8	99.6	5495	6	AX027812	AX027812 Sequence
21	716.8	99.6	6046	6	BD261826	BD261826 Recombina
22	716.8	99.6	6046	6	AX027823	AX027823 Sequence
23	716.8	99.6	6269	6	BD261820	BD261820 Recombina
24	716.8	99.6	6269	6	AX027817	AX027817 Sequence
25	716.8	99.6	6299	6	BD261821	BD261821 Recombina
26	716.8	99.6	6299	6	AX027818	AX027818 Sequence
27	716.8	99.6	6301	6	BD261817	BD261817 Recombina
28	716.8	99.6	6301	6	AX027814	AX027814 Sequence
29	716.8	99.6	39675	1	U00096_46	Continuation (47 o
30	716.8	99.6	110000	2	EC074221_1	Continuation (2 of
31	716.8	99.6	338534	1	EC0UW93_	U14003 Escherichia
32	716.8	99.6	720	6	AX087929	AX087929 Sequence
33	705.6	98.0	169764	1	AP002569	AP002569 Escherich
34	705.2	97.9	11599	1	AE005669	AE005669 Escherich
35	699.2	97.1	10535	1	AE015447	AE015447 Shigella
36	699.2	97.1	225944	1	AE016993	AE016993 Shigella
37	688	95.6	86898	1	AE016772	AE016772 Escherich
38	598.4	83.1	21405	1	AE008915	AE008915 Salmonell
39	587.2	81.6	88037	1	AL627284	AL627284 Salmonell
40	587.2	81.6	284233	1	AE016849	AE016849 Salmonell
41	558.4	77.6	732	6	AR384902	AR384902 Sequence
42	527.8	73.3	720	1	KSPNPGNE	X87881 Klebsiella
43	507.2	70.4	110000	1	EX950851_07	Continuation (8 of
44	480	66.7	11093	1	AE013977	AE013977 Yersinia
45	480	66.7	110000	1	EX936398_06	Continuation (7 of

ALIGNMENTS

RESULT 1	BD187724	BD187724	720 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	A method of producing a cytosine nucleoside compound.					
DEFINITION	BD187724	BD187724				
ACCESSION	BD187724.1	GI:32997463				
VERSION	JP 2003018997-A/3.					
KEYWORDS	Escherichia coli					
SOURCE	Escherichia coli					
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.					
REFERENCE	1 (bases 1 to 720)					
AUTHORS	Araki,T., Ikeda,I., Matoishi,K., Abe,R., Oikawa,T., Matsuba,Y., Nagahara,K., Fukui,Y. and Ishibashi,H.					
TITLE	A method of producing a cytosine nucleoside compound					
JOURNAL	Patent: JP 2003018997-A 3 21-JAN-2003;					
COMMENT	MITSUI CHEMICALS INC					
	OS	Escherichia coli				
	PN	JP 2003018997-A/3				
	PD	21-JAN-2003				
	PF	01-MAY-2002	JP 2002129867			
	PI	TADASHI ARAKI, ICHIRO IKEDA, KAORI MATOISHI, REIKO ABE, TOSHIHIRO				
	PI	OIKAWA,				
	PI	YASUKO MATSUBA, KIYOTERU NAGAHARA, YASUSHI FUKUIRI, HIROKI				
	ISHIBASHI					
	PC	C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/10				
	PC	C12P19/40, C12N15/00,				
	PC	C12N5/00				
	CC	A method of producing a cytosine nucleoside compound FH Key				
		Location/Qualifiers				
	FT	source	1..720			
	FT	Location/Qualifiers				
		1..720				
		/organism="Escherichia coli"				
		/mol_type="genomic DNA"				
		/db_xref="taxon:562"				
FEATURES						
source						

Query Match 99.6%; Score 716.8; DB 6; Length 720;
Best Local Similarity 99.7%; Pred. No. 1.6e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE 1
AUTHORS Tischer, W., Ihlenfeldt, H.G., Barzu, O., Sakamoto, H., Pistotnik, E.,
Marliere, P. and Pochet, S.
TITLE Enzymatic synthesis of deoxyribonucleosides
JOURNAL Patent: WO 0114566-A 3 01-MAR-2001;
Roche Diagnostics GmbH (DE) ; INSTITUT PASTEUR (FR) ; Pharma-
Waldhof GmbH & Co. KG (DE)

FEATURES
source Location/Qualifiers
1..720
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
CDS
1..720
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC34665.1"
/translation="MATPHINAEAGDPAADVLMGDPURAKYIAETFLFEDAREVNNVR
GMLGFTGYGRKISVMGHGIGPSCSYTKELITDFGVKKIIRVSGGAVLPHVKLR
DVVIGMGACTSKVNRIRFKDHPAAIADPMVRNVAADAALGIDARVGNLFADLRF
YSPDGEMFDVMEKYGILGVEMEAAGIYGVAAEFGAKALTICTVSDHIRTHTQTTAAER
QTTFNDMIKIALESVLLGDKE"

ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 720;
Best Local Similarity 99.7%; Pred. No. 1.6e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 60
DB 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 60
QY 61 GCGGACCCGCTCGCTGCGAAGTATATTCGTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
DB 61 GCGGACCCGCTCGCTGCGAAGTATATTCGTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
QY 121 AACACGTTCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
DB 121 AACACGTTCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
QY 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 240
DB 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 240
QY 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300
DB 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300
QY 301 CTGCGCGACGTGTTATCGGTATGGGTGCGTCCGACCGGATTCCAAAGTTAAACCGATCCGT 360
DB 301 CTGCGCGACGTGTTATCGGTATGGGTGCGTCCGACCGGATTCCAAAGTTAAACCGATCCGT 360
QY 361 TTTAAAGACCATGACTTTGCGCGTATCGGTGACTTCGACATGGTGCCTGTAACCGAGTAGAT 420
DB 361 TTTAAAGACCATGACTTTGCGCGTATCGGTGACTTCGACATGGTGCCTGTAACCGAGTAGAT 420
QY 421 GCAGCTAAAGCACTGGGTATTCGCTCGCGTGGGTAACTGTTCTCGCGCTGACCTGTTTC 480
DB 421 GCAGCTAAAGCACTGGGTATTCGCTCGCGTGGGTAACTGTTCTCGCGCTGACCTGTTTC 480
QY 481 TACTCTCCGGACCGCGAAATGTTCCACGTGATGGAAATACGGCAATTCCTGGCGTGAA 540
DB 481 TACTCTCCGGACCGCGAAATGTTCCACGTGATGGAAATACGGCAATTCCTGGCGTGAA 540
QY 541 ATGGAAGCGGCTGGTATCTACCGGCTCGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600
DB 541 ATGGAAGCGGCTGGTATCTACCGGCTCGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600
QY 601 TGCACCGGTATCTGACCATCCGCACTACCGAGCAGACCACTGCGCGCTGAGCGTCAGACT 660

Db 601 TGCACCGGTATCTGACCATCCGCACTCACAGCAGACCACTGCGCGTCAGACT 660
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720

RESULT 4
AX590437
LOCUS AX590437 720 bp DNA linear PAT 27-JAN-2003
DEFINITION Sequence 3 from Patent EP1254959.
ACCESSION AX590437
VERSION AX590437.1 GI:27949070
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE 1
AUTHORS Araki, T., Ikeda, I., Matoishi, K., Abe, R., Oikawa, T., Matsuba, Y.,
Ishibashi, H., Nagahara, K. and Fukui, Y.
TITLE Method for producing cytosine nucleoside compounds
JOURNAL Patent: EP 1254959-A 3 06-NOV-2002;
MITSUI CHEMICALS, INC. (JP)

FEATURES
source Location/Qualifiers
1..720
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"

ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 720;
Best Local Similarity 99.7%; Pred. No. 1.6e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 60
DB 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 60
QY 61 GCGGACCCGCTCGCTGCGAAGTATATTCGTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
DB 61 GCGGACCCGCTCGCTGCGAAGTATATTCGTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
QY 121 AACACGTTCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
DB 121 AACACGTTCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
QY 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 240
DB 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCACCAGT 240
QY 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300
DB 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300
QY 301 CTGCGCGACGTGTTATCGGTATGGGTGCGTCCGACCGGATTCCAAAGTTAAACCGATCCGT 360
DB 301 CTGCGCGACGTGTTATCGGTATGGGTGCGTCCGACCGGATTCCAAAGTTAAACCGATCCGT 360
QY 361 TTTAAAGACCATGACTTTGCGCGTATCGGTGACTTCGACATGGTGCCTGTAACCGAGTAGAT 420
DB 361 TTTAAAGACCATGACTTTGCGCGTATCGGTGACTTCGACATGGTGCCTGTAACCGAGTAGAT 420
QY 421 GCAGCTAAAGCACTGGGTATTCGCTCGCGTGGGTAACTGTTCTCGCGCTGACCTGTTTC 480
DB 421 GCAGCTAAAGCACTGGGTATTCGCTCGCGTGGGTAACTGTTCTCGCGCTGACCTGTTTC 480
QY 481 TACTCTCCGGACCGCGAAATGTTCCACGTGATGGAAATACGGCAATTCCTGGCGTGAA 540
DB 481 TACTCTCCGGACCGCGAAATGTTCCACGTGATGGAAATACGGCAATTCCTGGCGTGAA 540
QY 541 ATGGAAGCGGCTGGTATCTACCGGCTCGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600
DB 541 ATGGAAGCGGCTGGTATCTACCGGCTCGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600
QY 601 TGCACCGGTATCTGACCATCCGCACTACCGAGCAGACCACTGCGCGCTGAGCGTCAGACT 660

```
|||||
541 ATGGAAGCGGTGTAATCTACGGCGTCGTCAGAAATTTGGCGGAAGCCCTGACCATC 600
|||||
601 TGCACCGTATCTGACCAATCCGCACTACGAGCAGACCACTGCGCGTAGCGTCAGACT 660
|||||
601 TGCACCGTATCTGACCAATCCGCACTACGAGCAGACCACTGCGCGTAGCGTCAGACT 660
|||||
661 ACCTTACAGCATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGGATAAAGAGTAA 720
|||||
661 ACCTTACAGCATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGGATAAAGAGTAA 720
|||||

RESULT 5
ECOPNP 1249 bp DNA linear BCT 26-APR-1993
LOCUS E.coli purine nucleoside phosphorylase (deod) gene, complete cds.
DEFINITION M60917
ACCESSION M60917.1 GI:147308
VERSION 1
KEYWORDS purine nucleoside phosphorylase.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE Hersfield,M.S., Chaffee,S., Koro-Johnson,L., Mary,A., Smith,A.A.
AUTHORS (bases 1 to 1249)
TITLE Use of site-directed mutagenesis to enhance the epitope-shielding
effect of covalent modification of proteins with polyethylene
glycol
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (16), 7185-7189 (1991)
MEDLINE 91334430
PUBMED 1714590
COMMENT Original source text: Escherichia coli (strain K-12) DNA.
FEATURES
    source
        1..1249
            Location/Qualifiers
                /organism="Escherichia coli"
                /mol_type="genomic DNA"
                /strain="K-12"
                /db_xref="taxon:562"
                111..115
                    /note="putative"
                123..142
                    /gene="deod"
                123..142
                    /gene="deod"
                123..142
                    /gene="deod"
                /EC_number="2.4.2.1"
                /note="putative"
                /codon_start=1
                /transl_table=11
                /product="purine nucleoside phosphorylase"
                /protein_id="AAA24401.1"
                /db_xref="GI:147309"
                /translation="MATPHINAEMGDFADVLMGDFLRKRYIAETFLDAREVNNVR
                GMLGFTGYGRKISVNGHMGIPSCSIYTKELITDFGVKKIIRVSGCGLVPHVKLR
                DVIWGMCACTKVNIRFKDHPAAIDFMDVNAADAAKALGIDARVNLFSADLF
                YSPGEMFDVMEKYGLIGLVEAAGIYVAAEFKALGTICTVSDHIRTHEQTAAER
                OTTFNDIKIALESVLGDKE"
                874..901
                    /note="putative"
                /function="terminator for deod transcription"

    stem_loop
        99.6%; Score 716.8; DB 1; Length 1249;
        Best Local Similarity 99.7%; Pred. No.1.7e-181;
        Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

    1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 60
    123 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 182
    61 GGCACCCGCTGCGTGCAGAGTATATTCCTGGAACATTTTCCTTGAAGATGCCCGTGAAGTG 120
    183 GGCACCCGCTGCGTGCAGAGTATATTCCTGGAACATTTTCCTTGAAGATGCCCGTGAAGTG 242
```

```
QY 121 AACAACTTCCGCGTATGCTGGCGTTTACCGGTACTTTACAAAGCGCGCAAAATTTTCGGTA 180
Db 243 AACAACTTCCGCGTATGCTGGCGTTTACCGGTACTTTACAAAGCGCGCAAAATTTTCGGTA 302
QY 181 ATGGGTACACGGTGTGGTATCCCGTCTCTGCTCCATCTACACCAAGAAGTATGATCACCGAT 240
Db 303 ATGGGTACACGGTATGGTATCCCGTCTCTGCTCCATCTACACCAAGAAGTATGATCACCGAT 362
QY 241 TTGCGGTGAAGAAAAATATCCGCGTGGGTTCCTGTGGCGCAGTTTCTGCCGACGTAATAA 300
Db 363 TTGCGGTGAAGAAAAATATCCGCGTGGGTTCCTGTGGCGCAGTTTCTGCCGACGTAATAA 422
QY 301 CTGCGGACGCTCGTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAAACCCATCCGT 360
Db 423 CTGCGGACGCTCGTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAAACCCATCCGT 482
QY 361 TTTAAAGACCATGACTTTTCCGCTATCGCTGACTTCGACATGGTGGGTAAACGAGTAGAT 420
Db 483 TTTAAAGACCATGACTTTTCCGCTATCGCTGACTTCGACATGGTGGGTAAACGAGTAGAT 542
QY 421 GCAGCTTAAGCACTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTT 480
Db 543 GCAGCTTAAGCACTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTT 602
QY 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAATAATACGCGCATTTCTCGCGCTGGAA 540
Db 603 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAATAATACGCGCATTTCTCGCGCTGGAA 662
QY 541 ATGGAAGCGCTGCTATCTACCGCGTGGTGCAGAAATTTTGGCGCGAAAGCCCTGACCATC 600
Db 663 ATGGAAGCGCTGCTATCTACCGCGTGGTGCAGAAATTTTGGCGCGAAAGCCCTGACCATC 722
QY 601 TGACCGTATCTGACCACTCCGCACTCAGCAGCAGACCACTGCGCGTAGCGTCAGACT 660
Db 723 TGACCGTATCTGACCACTCCGCACTCAGCAGCAGACCACTGCGCGTAGCGTCAGACT 782
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 783 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 842

BD261823 3031 bp DNA linear PAT 17-JUL-2003
Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof.
BD261823
VERSION BD261823.1 GI:33071591
KEYWORDS JP 2002533126-A/12.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 3031)
AUTHORS Basetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL Patent: JP 2002533126-A 12 08-OCT-2002;
NORPHARMA SPA
OS Artificial Sequence
PN JP 2002533126-A/12
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BASTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI
ORSINI,
PI GIANCARLO TONON,GABRIELE ZUFFI
PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC
Description of Artificial Sequence: udp and deod cloned into CC
pgm746
CC without upstream ptac promoter
FH Key Location/Qualifiers
FT source 1..3031
```

FT /organism='Artificial Sequence'

FEATURES

source

Location/Qualifiers

1..3031

/organism="synthetic construct"

/mol_type="genomic DNA"

/db_xref="taxon:32630"

ORIGIN

Query Match 99.6%; Score 716.8; DB 6; Length 3031;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 60
Db 101 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 160
Qy 61 GGGACCCGCTGCGTGGCGAAGTATATTCCTGAACCTTTCCTTGAAGATGCCCGTGAAGTG 120
Db 161 GGGACCCGCTGCGTGGCGAAGTATATTCCTGAACCTTTCCTTGAAGATGCCCGTGAAGTG 220
Qy 121 AACAACTGTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA 180
Db 221 AACAACTGTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA 280
Qy 181 ATGGGTACCGGTGTTGGTATCCGTCCTGCTGCTCAATACAAAGAACTGATCACCGAT 240
Db 281 ATGGGTACCGGTGTTGGTATCCGTCCTGCTGCTCAATACAAAGAACTGATCACCGAT 340
Qy 241 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCCGACGTAATA 300
Db 341 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCCGACGTAATA 400
Qy 301 CTGCGCAGCTGCTTATCGGTATGGGTGCTGCACCGGATTCCTGCTGCTGCTGCTGCTGCT 360
Db 401 CTGCGCAGCTGCTTATCGGTATGGGTGCTGCACCGGATTCCTGCTGCTGCTGCTGCTGCT 460
Qy 361 TTTAAAGACCATGATTTTGGCGTATCGCTGATGCGATGCGATGCGATGCGATGCGATGCGAT 420
Db 461 TTTAAAGACCATGATTTTGGCGTATCGCTGATGCGATGCGATGCGATGCGATGCGATGCGAT 520
Qy 421 GCAGCTAAGCACCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTTTC 480
Db 521 GCAGCTAAGCACCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTTTC 580
Qy 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTCCTCGCGTGGAA 540
Db 581 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTCCTCGCGTGGAA 640
Qy 541 ATGGAAGCGGTGTTATCTACGGGTGCTGCGATGCGATGCGATGCGATGCGATGCGATGCGAT 600
Db 641 ATGGAAGCGGTGTTATCTACGGGTGCTGCGATGCGATGCGATGCGATGCGATGCGATGCGAT 700
Qy 601 TGCACCGTATCTGACCATCCGACCTACAGAGCAGACCACTGCGCGTGAACCTGAGCTGAGCT 660
Db 701 TGCACCGTATCTGACCATCCGACCTACAGAGCAGACCACTGCGCGTGAACCTGAGCTGAGCT 760
Qy 661 ACCTTCAACGACATGATCAAAATCGCAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 761 ACCTTCAACGACATGATCAAAATCGCAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 820

RESULT 7
AX027820 3031 bp DNA linear PAT 16-SEP-2000
LOCUS
DEFINITION Sequence 12 from Patent WO0039307.
ACCESSION AX027820
VERSION AX027820.1 GI:10188664
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.

REFERENCE 1
AUTHORS Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.

Recombinant bacterial strains for the production of natural

Journal

Patent: WO 0039307-A 12 06-JUL-2000;

BESTETTI GIUSEPPINA (IT); CALI SIMONA (IT); NORPHARMA SPA (IT);

ORSINI GAETANO (IT); TONON GIANCARLO (IT); ZUFFI GABRIELE (IT);

GHISOTTI DANIELA (IT)

FEATURES

source

Location/Qualifiers

1..3031

/organism="synthetic construct"

/mol_type="unassigned DNA"

/db_xref="taxon:32630"

/note="udp and deod cloned into pGM746 without upstream

ptac promoter"

ORIGIN

Query Match 99.6%; Score 716.8; DB 6; Length 3031;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 60
Db 101 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 160
Qy 61 GGGACCCGCTGCGTGGCGAAGTATATTCCTGAACCTTTCCTTGAAGATGCCCGTGAAGTG 120
Db 161 GGGACCCGCTGCGTGGCGAAGTATATTCCTGAACCTTTCCTTGAAGATGCCCGTGAAGTG 220
Qy 121 AACAACTGTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA 180
Db 221 AACAACTGTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA 280
Qy 181 ATGGGTACCGGTGTTGGTATCCGTCCTGCTGCTCAATACAAAGAACTGATCACCGAT 240
Db 281 ATGGGTACCGGTGTTGGTATCCGTCCTGCTGCTCAATACAAAGAACTGATCACCGAT 340
Qy 241 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCCGACGTAATA 300
Db 341 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCCGACGTAATA 400
Qy 301 CTGCGCAGCTGCTTATCGGTATGGGTGCTGCACCGGATTCCTGCTGCTGCTGCTGCTGCT 360
Db 401 CTGCGCAGCTGCTTATCGGTATGGGTGCTGCACCGGATTCCTGCTGCTGCTGCTGCTGCT 460
Qy 361 TTTAAAGACCATGATTTTGGCGTATCGCTGATGCGATGCGATGCGATGCGATGCGATGCGAT 420
Db 461 TTTAAAGACCATGATTTTGGCGTATCGCTGATGCGATGCGATGCGATGCGATGCGATGCGAT 520
Qy 421 GCAGCTAAGCACCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTTTC 480
Db 521 GCAGCTAAGCACCTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTTTC 580
Qy 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTCCTCGCGTGGAA 540
Db 581 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTCCTCGCGTGGAA 640
Qy 541 ATGGAAGCGGTGTTATCTACGGGTGCTGCGATGCGATGCGATGCGATGCGATGCGATGCGAT 600
Db 641 ATGGAAGCGGTGTTATCTACGGGTGCTGCGATGCGATGCGATGCGATGCGATGCGATGCGAT 700
Qy 601 TGCACCGTATCTGACCATCCGACCTACAGAGCAGACCACTGCGCGTGAACCTGAGCTGAGCT 660
Db 701 TGCACCGTATCTGACCATCCGACCTACAGAGCAGACCACTGCGCGTGAACCTGAGCTGAGCT 760
Qy 661 ACCTTCAACGACATGATCAAAATCGCAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db 761 ACCTTCAACGACATGATCAAAATCGCAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 820

RESULT 8

BD261824

LOCUS

DEFINITION

Recombinant bacterial strains for the production of natural

nucleosides and modified analogues thereof.

3128 bp DNA linear

PAT 17-JUL-2003

```
ACCESSION BD261824
VERSION BD261824.1 GI:33071592
KEYWORDS JP 2002533126-A/13.
SOURCE synthetic construct
ORGANISM synthetic construct; artificial sequences.
REFERENCE 1 (bases 1 to 3128)
AUTHORS Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL Patent: JP 2002533126-A 13 08-OCT-2002;
NORPHARMA SPA
COMMENT OS Artificial Sequence
PN JP 2002533126-A/13
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI
ORSINI,
PI GIANCARLO TONON,GABRIELE ZUFFI
PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC
Description of Artificial Sequence: deoD cloned downstream CC
ptac promoter
FH Key Location/Qualifiers
FT source 1..3128
FT /organism='Artificial Sequence'.
FEATURES
source Location/Qualifiers
1..3128
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 3128;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTGATGCCA 60
DB 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTGATGCCA 257
QY 61 GGCACCCGCTGCTGCGAAGTATATTGCTGAAACTTTTCCCTTGAAGATGCCCGTGAAGTG 120
DB 258 GGCACCCGCTGCTGCGAAGTATATTGCTGAAACTTTTCCCTTGAAGATGCCCGTGAAGTG 317
QY 121 AACAACTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCCGCAAAATTTCCGTA 180
DB 318 AACAACTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCCGCAAAATTTCCGTA 377
QY 181 ATGGGTACCGGTATGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
DB 378 ATGGGTACCGGTATGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 437
QY 241 TTCCGCGTGAAGAAAATTAATCCGCGTGGGTTCTGTGGCGAGTTCTGCCGACGTAAAA 300
DB 438 TTCCGCGTGAAGAAAATTAATCCGCGTGGGTTCTGTGGCGAGTTCTGCCGACGTAAAA 497
QY 301 CTGGCGACGTCGTATCGGTATGGGTGCGTGCACCGATTCCAAAGTTAACCGCATCCGT 360
DB 498 CTGGCGACGTCGTATCGGTATGGGTGCGTGCACCGATTCCAAAGTTAACCGCATCCGT 557
QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGCGAGTAGAT 420
DB 558 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGCGAGTAGAT 617
QY 421 GCAGCTAAAGCAGTGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTC 480
DB 618 GCAGCTAAAGCAGTGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTC 677
QY 481 TACTCTCGGACGCGGAAATTTGAGCTTCGAGTGTGAAAATACGCAATTCGCGGTGAA 540
DB 678 TACTCTCGGACGCGGAAATTTGAGCTTCGAGTGTGAAAATACGCAATTCGCGGTGAA 737
```

```
QY 541 ATGGAAGCGCTGCTATCTACGCGCTGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 600
DB 738 ATGGAAGCGCTGCTATCTACGCGCTGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 797
QY 601 TGCACCGTATCTGACACATCCGCACTCAGCAGACACCACTGCCGCTGAGCGTCAGACT 660
DB 798 TGCACCGTATCTGACACATCCGCACTCAGCAGACACCACTGCCGCTGAGCGTCAGACT 857
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCGGTTCTGCTGGCGGATAAAGAGTAA 720
DB 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCGGTTCTGCTGGCGGATAAAGAGTAA 917
RESULT 9
AX027821 3128 bp DNA linear PAT 16-SEP-2000
LOCUS Sequence 13 from Patent WO0039307.
DEFINITION AX027821
ACCESSION AX027821
VERSION AX027821.1 GI:10188665
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL Patent: WO 0039307-A 13 06-JUL-2000;
BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ;
ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
GHISOTTI DANIELA (IT)
FEATURES
Location/Qualifiers
source 1..3128
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="deoD cloned downstream ptac promoter"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 3128;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTGATGCCA 60
DB 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTTGATGCCA 257
QY 61 GGCACCCGCTGCTGCGAAGTATATTGCTGAAACTTTTCCCTTGAAGATGCCCGTGAAGTG 120
DB 258 GGCACCCGCTGCTGCGAAGTATATTGCTGAAACTTTTCCCTTGAAGATGCCCGTGAAGTG 317
QY 121 AACAACTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCCGCAAAATTTCCGTA 180
DB 318 AACAACTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCCGCAAAATTTCCGTA 377
QY 181 ATGGGTACCGGTATGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
DB 378 ATGGGTACCGGTATGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 437
QY 241 TTCCGCGTGAAGAAAATTAATCCGCGTGGGTTCTGTGGCGAGTTCTGCCGACGTAAAA 300
DB 438 TTCCGCGTGAAGAAAATTAATCCGCGTGGGTTCTGTGGCGAGTTCTGCCGACGTAAAA 497
QY 301 CTGGCGACGTCGTATCGGTATGGGTGCGTGCACCGATTCCAAAGTTAACCGCATCCGT 360
DB 498 CTGGCGACGTCGTATCGGTATGGGTGCGTGCACCGATTCCAAAGTTAACCGCATCCGT 557
QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGCGAGTAGAT 420
DB 558 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGCGAGTAGAT 617
QY 421 GCAGCTAAAGCAGTGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTC 480
```

```

Db      618 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCGCGTGACCTGTTTC 677
Qy      481 TACTCTCGGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCAATTCCTCGCGTGGAA 540
Db      678 TACTCTCGGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCAATTCCTCGCGTGGAA 737
Qy      541 ATGAAGCGGTGTTATCTACGGGTGCTCGCTGAGAAATTTGGGGGGAAGCCCTGACCATC 600
Db      738 ATGAAGCGGTGTTATCTACGGGTGCTCGCTGAGAAATTTGGGGGGAAGCCCTGACCATC 797
Qy      601 TGCACCGTATCTGACCATCCGCACTCACGAGCAGACCACTGCGCGTGAGCGTCAGACT 660
Db      798 TGCACCGTATCTGACCATCCGCACTCACGAGCAGACCACTGCGCGTGAGCGTCAGACT 857
Qy      661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db      858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 917

RESULT 10
BD261814
LOCUS      BD261814          3383 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Recombinant bacterial strains for the production of natural
ACCESSION  BD261814
VERSION     JP 2002533126-A/3.
KEYWORDS   JP 2002533126-A/3.
SOURCE     JP 2002533126-A/3.
ORGANISM   JP 2002533126-A/3.
REFERENCE  1 (bases 1 to 3383)
AUTHORS    Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE      Recombinant bacterial strains for the production of natural
           nucleosides and modified analogues thereof
JOURNAL    Patent: JP 2002533126-A 3 08-OCT-2002;
           NORPHARMA SPA
COMMENT    OS Artificial Sequence
           PN JP 2002533126-A/3
           PD 08-OCT-2002
           PR 23-DEC-1999 JP 2000591198
           PR 23-DEC-1998 IT MI 98A002792
           PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI
           ORSINI,
           PI GIANCARLO TONON,GABRIELE ZUFFI
           PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC
           Description of Artificial Sequence: Plasmid
           CC deod
           FH Key
           FT gene
           Location/Qualifiers
           1. .3383
           /organism="synthetic construct"
           /mol_type="genomic DNA"
           /db_xref="taxon:32630"

FEATURES
source
Query Match 99.6%; Score 716.8; DB 6; Length 3383;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

ORIGIN
1 ATGCTACCCACACATTAATGCGAAATGGCGGATTCGCTGACGTAGTTTGGATGCCA 60
240 ATGGCTACCCACACATTAATGCGAAATGGCGGATTCGCTGACGTAGTTTGGATGCCA 299
61 GCGACCGCTGCGTGGAGTATATGCTGAACCTTCCCTTGAAGATGCCCGTGAAGTG 120
300 GCGACCGCTGCGTGGAGTATATGCTGAACCTTCCCTTGAAGATGCCCGTGAAGTG 359
121 AACCAACGTTCCGCGTATGCTCGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
360 AACCAACGTTCCGCGTATGCTCGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 419
181 ATGGGTCAACCGTGTGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT 240

```

```

Db      420 ATGGGTCAACCGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT 479
Qy      241 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCCTGTGTGGCGCAGTTCTGCCCGCACGTAATA 300
Db      480 TTCCGCGTGAAGAAATTTATCCGCGTGGGTTCCTGTGTGGCGCAGTTCTGCCCGCACGTAATA 539
Qy      301 CTGCGCAACGTCGTTATCGGTATGGGTGCTGCAACCGATTCACAAAGTTAACCGCATCCGT 360
Db      540 CTGCGCAACGTCGTTATCGGTATGGGTGCTGCAACCGATTCACAAAGTTAACCGCATCCGT 599
Qy      361 TTTAAAGACCATGATTTTGGCGCTATCGCTGACTTTCGACATGGTGGCTTAACCGCATAGAT 420
Db      600 TTTAAAGACCATGATTTTGGCGCTATCGCTGACTTTCGACATGGTGGCTTAACCGCATAGAT 659
Qy      421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
Db      660 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 719
Qy      481 TACTCTCCGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCAATTCCTCGCGTGGAA 540
Db      720 TACTCTCCGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCAATTCCTCGCGTGGAA 779
Qy      541 ATGAAGCGGTGTTATCTACGGGTGCTGTCAGCAATTTGGCGGGAAGCCCTGACCATC 600
Db      780 ATGAAGCGGTGTTATCTACGGGTGCTGTCAGCAATTTGGCGGGAAGCCCTGACCATC 839
Qy      601 TGCACCGTATCTGACCATCCGCACTCACGAGCAGACCACTGCGCGTGAGCGTCAGACT 660
Db      840 TGCACCGTATCTGACCATCCGCACTCACGAGCAGACCACTGCGCGTGAGCGTCAGACT 899
Qy      661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db      900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 959

RESULT 11
AX027811
LOCUS      AX027811          3383 bp      DNA      linear      PAT 16-SEP-2000
DEFINITION Sequence 3 from Patent WO0039307.
ACCESSION  AX027811
VERSION     AX027811.1 GI:10188655
KEYWORDS   synthetic construct
           other sequences; artificial sequences.
SOURCE     synthetic construct
           other sequences; artificial sequences.
REFERENCE  1
AUTHORS    Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE      Recombinant bacterial strains for the production of natural
           nucleosides and modified analogues thereof
JOURNAL    Patent: WO 0039307-A 3 06-JUL-2000;
           BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ;
           ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
           GHISOTTI DANIELA (IT)
           Location/Qualifiers
           1. .3383
           /organism="synthetic construct"
           /mol_type="unassigned DNA"
           /db_xref="taxon:32630"
           /note="Plasmid"

ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 3383;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 ATGCTACCCACACATTAATGCGAAATGGCGGATTCGCTGACGTAGTTTGGATGCCA 60
240 ATGGCTACCCACACATTAATGCGAAATGGCGGATTCGCTGACGTAGTTTGGATGCCA 299
61 GCGACCGCTGCGTGGAGTATATGCTGAACCTTCCCTTGAAGATGCCCGTGAAGTG 120
300 GCGACCGCTGCGTGGAGTATATGCTGAACCTTCCCTTGAAGATGCCCGTGAAGTG 359

```

```
QY 121 AACAACTTCGCGGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 180
Db 360 AACAACTTCGCGGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 419
QY 181 ATGGGTACACCGTGTGGGTATCCCGTCTCTGCTTCCATCTACACCAAGAACTGATCACCGAT 240
Db 420 ATGGGTACACCGTGTGGGTATCCCGTCTCTGCTTCCATCTACACCAAGAACTGATCACCGAT 479
QY 241 TTCCGGCTGAAGAAATATTCGCGGTGGGTTCCTGTGGGAGTTCTGCGCGACGTAAATA 300
Db 480 TTCCGGCTGAAGAAATATTCGCGGTGGGTTCCTGTGGGAGTTCTGCGCGACGTAAATA 539
QY 301 CTGCGCACGCTCGTATCGGTATGGGTGCGCTGCACCGATTCCTCAAGTTTACCGATCCGT 360
Db 540 CTGCGCACGCTCGTATCGGTATGGGTGCGCTGCACCGATTCCTCAAGTTTACCGATCCGT 599
QY 361 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGCCTAACCGCAGTAGAT 420
Db 600 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGCCTAACCGCAGTAGAT 659
QY 421 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
Db 660 GCAGCTAAAGCAGCTGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 719
QY 481 TACTCTCCGACGCGGAAATGTTCCGAGCTGATGGAGAAATACCGCATTTCTCGCGTGGAA 540
Db 720 TACTCTCCGACGCGGAAATGTTCCGAGCTGATGGAGAAATACCGCATTTCTCGCGTGGAA 779
QY 541 ATGGAAGCGGTGCTGATCTACGCGGTGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db 780 ATGGAAGCGGTGCTGATCTACGCGGTGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 839
QY 601 TGCACCGTATCTGACCAATCCGACATCTACAGAGCAGACCACTGCGCGCTGAGCGTCAGACT 660
Db 840 TGCACCGTATCTGACCAATCCGACATCTACAGAGCAGACCACTGCGCGCTGAGCGTCAGACT 899
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 959

RESULT 12
BD261825 3934 bp DNA linear PAT 17-JUL-2003
LOCUS Recombinant bacterial strains for the production of natural
DEFINITION nucleosides and modified analogues thereof.
ACCESSION BD261825
VERSION 1 (bases 1 to 3934)
KEYWORDS JP 2002533126-A/14.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 3934)
AUTHORS Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL Patent: JP 2002533126-A 14 08-OCT-2002;
NORPHARMA SPA
COMMENT OS Artificial Sequence
PN JP 2002533126-A/14
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI
ORSINI,
PI GIANCARLO TONON,GABRIELE ZUFFI
PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC
Description of Artificial Sequence: udp and deod cloned CC
downstream ptac
CC promoter
FH Key Location/Qualifiers
FT source 1..3934
```

```
FEATURES             Location/Qualifiers
     source            1..3934
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"

ORIGIN
Query Match       99.6%; Score 716.8; DB 6; Length 3934;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY   1  ATGCGTACCCACACACATTAAATGAGAAATGGGCGATTTTCGCTGACGCTAGTTTGTGATGCCA 60
Db   198 ATGCGTACCCACACACATTAAATGAGAAATGGGCGATTTTCGCTGACGCTAGTTTGTGATGCCA 257
QY   61  GGCAGCCGCTGCGTGGAGATATATTTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 120
Db   258 GGCAGCCGCTGCGTGGAGATATATTTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 317
QY   121 AACAGGTTGCGGATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db   318 AACAGGTTGCGGATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 377
QY   181 ATGGGTACACGCTGTGGTATCCGCTCCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
Db   378 ATGGGTACACGCTGTGGTATCCGCTCCTGCTCCATCTACACCAAGAACTGATCACCGAT 437
QY   241 TTCCGGGTGAAGAAATATATCCGCGTGGGTTCCTGTGGCGCAGTTCTTGC CGCAGTAAAA 300
Db   438 TTCCGGGTGAAGAAATATATCCGCGTGGGTTCCTGTGGCGCAGTTCTTGC CGCAGTAAAA 497
QY   301 CTGGCGAGCTGCTTATCCGCTATGGGTATGGGTGCTGCACCGATTCCTCAAGTTTACCGCATCCGT 360
Db   498 CTGGCGAGCTGCTTATCCGCTATGGGTATGGGTGCTGCACCGATTCCTCAAGTTTACCGCATCCGT 557
QY   361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGCCTAAACGCGTAGAT 420
Db   558 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGCCTAAACGCGTAGAT 617
QY   421 GCAGCTAAAGCAGCTGGGTATGATGCTGCGGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
Db   618 GCAGCTAAAGCAGCTGGGTATGATGCTGCGGTGGGTAACTGTTCTCCGCTGACCTGTTTC 677
QY   481 TACTCTCCGACGCGGAAATGTTCCGAGCTGATGGAGAAATACGSCATTTCTCGCGTGGAA 540
Db   678 TACTCTCCGACGCGGAAATGTTCCGAGCTGATGGAGAAATACGSCATTTCTCGCGTGGAA 737
QY   541 ATGGAAGCGGCTGCTATCTACGCGGCTGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db   738 ATGGAAGCGGCTGCTATCTACGCGGCTGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 797
QY   601 TGACCGGTATCTGACCAATCCGACATCTACAGAGCAGACCACTGCGCGTGAAGCGTCAGACT 660
Db   798 TGACCGGTATCTGACCAATCCGACATCTACAGAGCAGACCACTGCGCGTGAAGCGTCAGACT 857
QY   661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db   858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 917

RESULT 13
AX027822
LOCUS AX027822 3934 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 14 from Patent WO0039307.
ACCESSION AX027822
VERSION AX027822.1 GI:10188666
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
```

TITLE Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof
JOURNAL Patent: WO 0039307-A 14 06-JUL-2000;
BESTETTI GIUSEPPINA (IT); CALI SIMONA (IT); NORPHARMA SPA (IT); ORSINI GAETANO (IT); TONON GIANCARLO (IT); ZUFFI GABRIELE (IT); GHISOTTI DANIELA (IT)
FEATURES Location/Qualifiers
source 1..3934
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="udp and deod cloned downstream ptac promoter"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 3934;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGGATGCCA 60
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGGATGCCA 257
Qy 61 GGGGACCCGCTGCGTGGAGATATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
Db 258 GGGGACCCGCTGCGTGGAGATATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 317
Qy 121 AACAACTGTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCGCAAAATTTCCGTA 180
Db 318 AACAACTGTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCGCAAAATTTCCGTA 377
Qy 181 ATGGGTACCGGTATGCTGGTATCCGCTGCTGCTCATCTACACAAAGAACTGATCACCAGT 240
Db 378 ATGGGTACCGGTATGCTGGTATCCGCTGCTGCTCATCTACACAAAGAACTGATCACCAGT 437
Qy 241 TTCCGCGTGAAGAAATATCCGCGTGGTTCCTGTGGCGAGTTCTGCCGACGTAAAA 300
Db 438 TTCCGCGTGAAGAAATATCCGCGTGGTTCCTGTGGCGAGTTCTGCCGACGTAAAA 497
Qy 301 CTGGCGACGCTGTTATCGGTATGGGTGCTGCGTGGTAACTTCCAAAGTTAACCGCATCCGT 360
Db 498 CTGGCGACGCTGTTATCGGTATGGGTGCTGCGTGGTAACTTCCAAAGTTAACCGCATCCGT 557
Qy 361 TTTAAAGACCATGACTTTGCGGTATCGCTGACTTCGACATGGTGGTAAAGCATAGAT 420
Db 558 TTTAAAGACCATGACTTTGCGGTATCGCTGACTTCGACATGGTGGTAAAGCATAGAT 617
Qy 421 GCAGCTAAAGCAGTGGTATGATGCTGCGTGGTAACTTCCGCTGACCTGTTTC 480
Db 618 GCAGCTAAAGCAGTGGTATGATGCTGCGTGGTAACTTCCGCTGACCTGTTTC 677
Qy 481 TACTCTCCGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGAA 540
Db 678 TACTCTCCGACGGCGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGAA 737
Qy 541 ATGAAGCGGTGATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db 738 ATGAAGCGGTGATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 797
Qy 601 TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCACTGCGCTGAGCTCAGACT 660
Db 798 TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCACTGCGCTGAGCTCAGACT 857
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 917

RESULT 14
BD261816 4189 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof.
ACCESSION BD261816

VERSION BD261816.1 GI:33071584
KEYWORDS JP 2002533126-A/5.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 4189)
AUTHORS Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE Recombinant bacterial strains for the production of natural nucleosides and modified analogues thereof
JOURNAL Patent: JP 2002533126-A 5 08-OCT-2002;
NORPHARMA SPA
COMMENT OS Artificial Sequence
PN JP 2002533126-A/5
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI ORSINI,
PI GIANCARLO TONON,GABRIELE ZUFFI
PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC Description of Artificial Sequence: Plasmid
CC udp
CC deod
FH Key Location/Qualifiers
FT gene (243)..(1021)
FT gene (1037)..(1766).
FEATURES Location/Qualifiers
source 1..4189
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 99.6%; Score 716.8; DB 6; Length 4189;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGGATGCCA 60
Db 1046 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGGATGCCA 1105
Qy 61 GGGGACCCGCTGCGTGGAGATATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
Db 1106 GGGGACCCGCTGCGTGGAGATATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 1165
Qy 121 AACAACTGTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCGCAAAATTTCCGTA 180
Db 1166 AACAACTGTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGGCGCAAAATTTCCGTA 1225
Qy 181 ATGGGTACCGGTATGGTATCCCGTCTGCTCATCTACACAAAGAACTGATCACCAGT 240
Db 1226 ATGGGTACCGGTATGGTATCCCGTCTGCTCATCTACACAAAGAACTGATCACCAGT 1285
Qy 241 TTCCGCGTGAAGAAATATCCGCGTGGTTCCTGTGGCGAGTTCTGCCGACGTAAAA 300
Db 1286 TTCCGCGTGAAGAAATATCCGCGTGGTTCCTGTGGCGAGTTCTGCCGACGTAAAA 1345
Qy 301 CTGGCGACCTGCTTATCGGTATGGGTGCTGCAACCGATTCCAAAGTTAACCGCATCCGT 360
Db 1346 CTGGCGACCTGCTTATCGGTATGGGTGCTGCAACCGATTCCAAAGTTAACCGCATCCGT 1405
Qy 361 TTTAAAGACCATGACTTTTGGCGGTATCGCTGACTTTCGACATGGTGGTAAAGCATAGAT 420
Db 1406 TTTAAAGACCATGACTTTTGGCGGTATCGCTGACTTTCGACATGGTGGTAAAGCATAGAT 1465
Qy 421 GCAGCTAAAGCAGTGGGTATGATGCTGCGGTGGGTAACTGTTCTCGCTGACCTGTTTC 480
Db 1466 GCAGCTAAAGCAGTGGGTATGATGCTGCGGTGGGTAACTGTTCTCGCTGACCTGTTTC 1525
Qy 481 TACTCTCCGACGGCGAAATGTTTCGAGCTGATGGAATAATACGGCATTTCTCGGGGTGAA 540
Db 1526 TACTCTCCGACGGCGAAATGTTTCGAGCTGATGGAATAATACGGCATTTCTCGGGGTGAA 1585

```
QY 541 ATGGAAGCGGTGTATCTACGGCTCGCTGCAGAAATTTGCGCGAAGCCCTGACCATC 600
    |||||
Db 1586 ATGGAAGCGGTGTATCTACGGCTCGCTGCAGAAATTTGCGCGAAGCCCTGACCATC 1645
    |||||
QY 601 TGCACCGTATCTGACCAATCCGCACTACAGAGCAGACCACTGCGCTGAGCGTCAGACT 660
    |||||
Db 1646 TGCACCGTATCTGACCAATCCGCACTACAGAGCAGACCACTGCGCTGAGCGTCAGACT 1705
    |||||
QY 661 ACCTTCAACGACATGATCAAAATCGCACTCGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
    |||||
Db 1706 ACCTTCAACGACATGATCAAAATCGCACTCGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 1765

RESULT 15
AX027813
LOCUS AX027813 4189 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 5 from Patent W00039307.
ACCESSION AX027813
VERSION AX027813.1 GI:10188657
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE
1 Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
  Recombinant bacterial strains for the production of natural
  nucleosides and modified analogues thereof
  Patent: WO 0039307-A 5 06-JUL-2000;
  BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ;
  ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
  GHISOTTI DANIELA (IT)
FEATURES
    source
    location/Qualifiers
    1..4189
    /organism="synthetic construct"
    /mol_type="unassigned DNA"
    /db_xref="taxon:32630"
    /note="Plasmid"
ORIGIN

Query Match 99.6%; Score 716.8; DB 6; Length 4189;
Best Local Similarity 99.7%; Pred. No. 1.8e-181;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGAGAAATGGCGATTTTCGCTGACGTAGTTTTCATGCCA 60
    |||||
Db 1046 ATGGCTACCCACACATTAATGAGAAATGGCGATTTTCGCTGACGTAGTTTTCATGCCA 1105
    |||||
QY 61 GGGGACCCGCTGGCTGCGAAGTATATTGCTCAAACTTTCTTGAAGATGCCCGTGAAGTG 120
    |||||
Db 1106 GGGGACCCGCTGGCTGCGAAGTATATTGCTCAAACTTTCTTGAAGATGCCCGTGAAGTG 1165
    |||||
QY 121 AACAACTTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGCCGCCAAAATTTCCGTA 180
    |||||
Db 1166 AACAACTTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGCCGCCAAAATTTCCGTA 1225
    |||||
QY 181 ATGGGTACCGGTGTGGTATCCCGTCCGTCCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
    |||||
Db 1226 ATGGGTACCGGTATGGGTATCCCGTCCGTCCTGCTCCATCTACACCAAGAACTGATCACCGAT 1285
    |||||
QY 241 TTGGCGGTGAAGAAAATTTATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGCACGTAATAA 300
    |||||
Db 1286 TTGGCGGTGAAGAAAATTTATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGCACGTAATAA 1345
    |||||
QY 301 CTGCGCGACGTGCTGTATCGGTATGGGTGCTTGCAACCGATTCCAAAGTTAACCGCATCCGT 360
    |||||
Db 1346 CTGCGCGACGTGCTGTATCGGTATGGGTGCTTGCAACCGATTCCAAAGTTAACCGCATCCGT 1405
    |||||
QY 361 TTTAAAGACCATGACTTTTGGCGTATCGCTGACTTCGACATGGTGCCTAACGCGAGTAGAT 420
    |||||
Db 1406 TTTAAAGACCATGACTTTTGGCGTATCGCTGACTTCGACATGGTGCCTAACGCGAGTAGAT 1465
    |||||
QY 421 GCAGCTAAAGCACTGGGTATTGATGCTCGGCTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
    |||||
```

```
Db 1466 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 1525
    |||||
QY 481 TACTCTCCGACCGCGGAAATTTGACGCTGATCGGAAATATACGGCATTTCTCGGCGTGGAA 540
    |||||
Db 1526 TACTCTCCGACCGCGGAAATTTGACGCTGATCGGAAATATACGGCATTTCTCGGCGTGGAA 1585
    |||||
QY 541 ATGGAAGCGGTGTATCTACGGCTCGCTGCAGAAATTTGGCGCGGAAAGCCCTGACCATC 600
    |||||
Db 1586 ATGGAAGCGGTGTATCTACGGCTCGCTGCAGAAATTTGGCGCGGAAAGCCCTGACCATC 1645
    |||||
QY 601 TGCACCGTATCTGACCAATCCGCACTACAGAGCAGACCACTGCGCTGAGCGTCAGACT 660
    |||||
Db 1646 TGCACCGTATCTGACCAATCCGCACTACAGAGCAGACCACTGCGCTGAGCGTCAGACT 1705
    |||||
QY 661 ACCTTCAACGACATGATCAAAATCGCACTCGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
    |||||
Db 1706 ACCTTCAACGACATGATCAAAATCGCACTCGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 1765
    |||||
```

Search completed: July 22, 2005, 21:18:03
Job time : 5511 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 21, 2005, 23:14:29 ; Search time 720 Seconds
(without alignments)
5919.741 Million cell updates/sec

Title: US-10-035-300A-1

Perfect score: 720

Sequence: 1 atggctaccacacattaa.....tgctggcgtaagaagtaa 720

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	720	100.0	720	8	Aad56237 Escherich
2	716.8	99.6	720	5	Aaf55440 Nucleotid
3	716.8	99.6	720	8	Aad54112 Escherich
4	716.8	99.6	720	12	Adl27841 E coli de
5	716.8	99.6	752	6	Ab151595 Purine nu
6	716.8	99.6	3383	3	Aaa51632 Plasmid p
7	716.8	99.6	4189	3	Aaa51634 Plasmid p
8	716.8	99.6	5013	9	Aca63355 E. coli D
9	716.8	99.6	5013	10	Aad59423 E.coli pu
10	716.8	99.6	5013	10	Ade86127 E. coli p
11	716.8	99.6	5495	3	Aaa51633 Plasmid p
12	716.8	99.6	6301	3	Aaa51635 Plasmid p
13	715.2	99.3	720	8	Aad56238 Escherich
14	713.6	99.1	720	5	Aaf55446 Nucleotid
15	639.8	88.9	3066	3	Aaa51641 pGM747 co
16	638.8	88.7	5303	3	Aaa51636 Plasmid p
17	636.8	88.4	3165	3	Aaa51642 pGM751 co
18	636.8	88.4	3979	3	Aaa51643 pGM800 wi
19	636.8	88.4	6117	3	Aaa51644 pGM807 wi
20	632.8	87.9	6373	3	Aaa51639 pGM795 co

21	628.8	87.3	6343	3	AAA51638	Aaa51638 Plasmid p
22	570	79.2	2807	5	AAS86794	Aas86794 DNA encod
23	558.8	77.6	1112	5	AAS89111	Aas89111 DNA encod
24	558.4	77.6	732	11	ACH95836	Ach95836 Klebsiell
25	533.6	74.1	1893	5	AAS88507	Aas88507 DNA encod
26	449.2	62.4	110000	10	ACF65385	Acf65385 Phototrab
27	449.2	62.4	110000	10	ACF67367_29	Continuation (30 o
28	448.8	62.3	717	10	ACF69888	Acf69888 Phototrab
29	438.4	60.9	783	5	AAS86792	Aas86792 DNA encod
30	420.8	58.4	795	10	ADF02086	Adf02086 Bacterial
31	403.6	56.1	6701	13	ADT05517	Adt05517 Haemophil
32	403.6	56.1	85814	13	ADT05644	Adt05644 Haemophil
33	402.6	55.9	110000	2	AAT42063_05	Continuation (6 of
34	330	45.8	5890	3	AAA51637	Aaa51637 Plasmid p
35	243.4	33.8	110000	6	ABA92787_5	Continuation (6 of
36	212.2	29.5	110000	6	ABA90521_09	Continuation (10 o
37	210.8	29.3	708	8	ACF73435	Acf73435 Staphyloc
38	210.8	29.3	5301	2	AAV74754	AAv74754 Staphyloc
39	206.6	28.7	2040	2	AAT77102	Aat77102 DNA encod
40	205.8	28.6	2678	5	AAS89114	Aas89114 DNA encod
41	205.4	28.5	899	13	ADRI2667	Adri2667 Bacillus
42	204.2	28.4	708	10	ABX06511	Abx06511 S. pneumo
43	204.2	28.4	711	2	AAV72439	AAv72439 S. pneumo
44	204.2	28.4	711	2	AAV72438	AAv72438 S. pneumo
45	204.2	28.4	714	13	ADR93332	Adr93332 Novel S.

ALIGNMENTS

RESULT 1	
AAD56237	
ID	AAD56237 standard; DNA; 720 BP.
AC	AAD56237;
XX	
DT	07-AUG-2003 (first entry)
DE	Escherichia coli mutant PNP enzyme (M65V) encoding DNA #SEQ ID 1.
XX	
KW	Purine nucleoside phosphorylase; mutant; tumour; PNP; enzyme; cancer;
KW	leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy;
KW	gene; ds.
XX	
OS	Escherichia coli.
XX	
XX	Synthetic.
XX	
FT	Key
CDS	Location/Qualifiers
FT	1..720
FT	/*tag= a
FT	/product= "Mutant PNP enzyme"
XX	
PN	WO2003035012-A2.
XX	
PD	01-MAY-2003.
XX	
PF	28-OCT-2002; 2002WO-US034626.
XX	
PR	26-OCT-2001; 2001US-00035300.
XX	
PA	(UABR-) UAB RES FOUND.
PA	(SOUR) SOUTHERN RES INST.
PA	(CORR) CORNELL RES FOUND INC.
XX	
PI	Ealick SE, Parker WB, Seerist JA, Sorscher EJ;
XX	
DR	WPI; 2003-421350/39.
DR	P-ESDB; AAE37217.
XX	
PT	Novel purified mutant purine cleaving enzyme having different biological
PT	activity than a wild-type purine cleaving enzyme, useful for treating a
XX	pathological condition characterized by abnormal cell growth.

PS Claim 10; Page 63-64; 67pp; English.

XX The invention relates to mutant purine nucleoside phosphorylase (PNP)

CC enzymes and nucleic acid molecules encoding such enzymes. These mutant

CC enzymes have greater biological activity than wild-type enzymes. PNP

CC enzymes are useful for treating a pathological condition characterised by

CC abnormal cell growth such as that occurring in cancer of the skin,

CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,

CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,

CC thyroid, testicle and uterus as well as other conditions characterised by

CC abnormal cell growth such as myeloid leukaemia, glioblastoma and

CC lymphoma. Sequences of the invention are useful for treating metastatic

CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or

CC to activate prodrugs in the treatment of cancer. They are useful for

CC treating virally infected cells. They are also useful in gene therapy.

CC The present sequence is Escherichia coli mutant PNP enzyme (M65V)

CC encoding DNA

XX

SQ Sequence 720 BP; 171 A; 190 C; 187 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 720; DB 8; Length 720;

Best Local Similarity 100.0%; Pred. No. 9.2e-218;

Matches 720; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAAATGAGAAATGGCGATTTCGCTGACGTAGTTTTCATGCCA 60

Db 1 ATGGCTACCCACACATTAAATGAGAAATGGCGATTTCGCTGACGTAGTTTTCATGCCA 60

Qy 61 GGCACCCGCTGCGTGGAGATATATTGCTGAAACTTTTCCTTGAAGATGCCGTTGAAGTG 120

Db 61 GGCACCCGCTGCGTGGAGATATATTGCTGAAACTTTTCCTTGAAGATGCCGTTGAAGTG 120

Qy 121 AACACGTTCCGCTATGCTGGCTTACCGGTACTTACAAAGCCGCAAAATTCGCTA 180

Db 121 AACACGTTCCGCTATGCTGGCTTACCGGTACTTACAAAGCCGCAAAATTCGCTA 180

Qy 181 ATGGGTCAACGCTGTTGGTATCCCTGCTCTGCTTACACCAAGAACATGATCACCGAT 240

Db 181 ATGGGTCAACGCTGTTGGTATCCCTGCTCTGCTTACACCAAGAACATGATCACCGAT 240

Qy 241 TTCGGCTGAAGAAATATATCCGCTGGTTCCTGTGGCGAGTTTCCTGGCGACGTTAAA 300

Db 241 TTCGGCTGAAGAAATATATCCGCTGGTTCCTGTGGCGAGTTTCCTGGCGACGTTAAA 300

Qy 301 CTGGCGACGCTGTTATCGGTATCGGTGCTGACCGATTCCTGCTGCTGCTGCTGCTGCT 360

Db 301 CTGGCGACGCTGTTATCGGTATCGGTGCTGACCGATTCCTGCTGCTGCTGCTGCTGCT 360

Qy 361 TTTAAAGACCATGACTTTGCGCTATCGCTGACTTCGACATGCTGCTGCTGCTGCTGCTG 420

Db 361 TTTAAAGACCATGACTTTGCGCTATCGCTGACTTCGACATGCTGCTGCTGCTGCTGCTG 420

Qy 421 GCAGCTAAAGCAGTGGTATTTGATGCTGCGTGGGTAACTGTTCTCGCTGACCTGTTTC 480

Db 421 GCAGCTAAAGCAGTGGTATTTGATGCTGCGTGGGTAACTGTTCTCGCTGACCTGTTTC 480

Qy 481 TACTCTCGGACGCGAAATGTTTCGAGCTGATGCAAAATACGCAATTCCTGCGGCTGAA 540

Db 481 TACTCTCGGACGCGAAATGTTTCGAGCTGATGCAAAATACGCAATTCCTGCGGCTGAA 540

Qy 541 ATGGAAGCGGCTGTTATCTACGCGTTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600

Db 541 ATGGAAGCGGCTGTTATCTACGCGTTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600

Qy 601 TGCAACGCTATGACACATCCGACATCAAGAGAGACCACTGCGCTGAGCGTGCAGACT 660

Db 601 TGCAACGCTATGACACATCCGACATCAAGAGAGACCACTGCGCTGAGCGTGCAGACT 660

Qy 661 ACCTTCAACGACATGATCAAAATCGCAATTCGCTGCGGCGATAAAGAGTAA 720

Db 661 ACCTTCAACGACATGATCAAAATCGCAATTCGCTGCGGCGATAAAGAGTAA 720

RESULT 2

AAF55440

ID AAF55440 standard; DNA; 720 BP.

XX

AC AAF55440;

XX

DT 29-MAY-2001 (first entry)

XX

DE Nucleotide sequence of a purine nucleoside phosphorylase.

XX

KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;

KW purine nucleoside phosphorylase; phosphopentose mutase;

KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;

KW deoxyribokinase; nucleoside 2-deoxyribosyltransferase; ss.

XX

OS Escherichia coli.

XX

FH Key Location/Qualifiers

FT CDS 1..720

FT /*tag= a

FT /product= "purine nucleoside phosphorylase"

XX

PN WO200114566-A2.

XX

PD 01-MAR-2001.

XX

PF 18-AUG-2000; 2000WO-EP008088.

XX

PR 20-AUG-1999; 99EP-00116425.

XX

PA (HOFF) ROCHE DIAGNOSTICS GMBH.

PA (INSP) INST PASTEUR.

PA (PHAR-) PHARMA-WALDHOF GMBH & CO KG.

XX

PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;

PI Marliere P, Pochet S;

XX

DR WPI: 2001-235026/24.

DR P-PSDB; AAB67584.

XX

PT In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting

PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside

PT and an inorganic phosphate.

XX

PS Disclosure; Page 44-45; 73pp; English.

XX

CC The present sequence encodes a purine nucleoside phosphorylase enzyme.

CC This enzyme is involved in the biosynthesis of deoxyribonucleosides, and

CC is used in the method of the invention. The specification describes a

CC method for the in vitro enzymatic synthesis of deoxyribonucleosides. The

CC method comprises reacting deoxyribose 1-phosphate and a nucleobase to

CC form a deoxyribonucleoside and an inorganic phosphate. Enzymes which may

CC be used in the method of the invention include thymidine phosphorylase,

CC purine nucleoside phosphorylase, phosphopentose mutase, phosphopentose

CC aldolase, fructose 1,6-diphosphate aldolase, deoxyribokinase, and

CC nucleoside 2-deoxyribosyltransferase

XX

SQ Sequence 720 BP; 172 A; 190 C; 187 G; 171 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 5; Length 720;

Best Local Similarity 99.7%; Pred. No. 9.6e-217;

Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAAATGAGAAATGGCGATTTCGCTGACGTAGTTTTCATGCCA 60

Db 1 ATGGCTACCCACACATTAAATGAGAAATGGCGATTTCGCTGACGTAGTTTTCATGCCA 60

Qy 61 GGCACCCGCTGCGTGGAGATATATTGCTGAAACTTTTCCTTGAAGATGCCGTTGAAGTG 120

Db 61 GGCACCCGCTGCGTGGAGATATATTGCTGAAACTTTTCCTTGAAGATGCCGTTGAAGTG 120

Qy 121 AACACGTTCCGCTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTCGCTA 180

Db 121 AACAAAGTTCGGGTATGCTGGGCTTACCGGTATCTTACAAAGCGCGCAAAATTTCCGTA 180
Qy 181 ATGGTCAAGCTGTTGGTATCCCGTCCCTGCTACACCAAGAACTGATCACCGAT 240
Db 181 ATGGTCAAGCTGTTGGTATCCCGTCCCTGCTACACCAAGAACTGATCACCGAT 240
Qy 241 TTGGCGTGAAGAAATTTATCCCGTGGGTCTGTCGGCGAGTTCTGCGCGACGTAAAA 300
Db 241 TTGGCGTGAAGAAATTTATCCCGTGGGTCTGTCGGCGAGTTCTGCGCGACGTAAAA 300
Qy 301 CTGGCGAGCTGTTATCGGTATGGTGGTCTGTCACCGGATTCAGAAATTTACCGCATCCGT 360
Db 301 CTGGCGAGCTGTTATCGGTATGGTGGTCTGTCACCGGATTCAGAAATTTACCGCATCCGT 360
Qy 361 TTTAAAGACCATGACTTTTCCCGCTATCGCTGACTTCGACATGGTGGTAAAGCGATAGAT 420
Db 361 TTTAAAGACCATGACTTTTCCCGCTATCGCTGACTTCGACATGGTGGTAAAGCGATAGAT 420
Qy 421 GCAGCTAAAGCACTGGGTATGATGCTCGGTGGGTAAACCTGTTCTCGGCTGACCTGTT 480
Db 421 GCAGCTAAAGCACTGGGTATGATGCTCGGTGGGTAAACCTGTTCTCGGCTGACCTGTT 480
Qy 481 TACTCTCCGGACGGCGAAATGTTGACCTGATGGAAATACGGCATTTCTGGCGTGGAA 540
Db 481 TACTCTCCGGACGGCGAAATGTTGACCTGATGGAAATACGGCATTTCTGGCGTGGAA 540
Qy 541 ATGAAGCGGTGGTATCTAGCGGTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db 541 ATGAAGCGGTGGTATCTAGCGGTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 601 TGCACCGTATGACCACTACCGCACTACGAGCAGACCACTGCGGTGAGCGTCAGACT 660
Db 601 TGCACCGTATGACCACTACCGCACTACGAGCAGACCACTGCGGTGAGCGTCAGACT 660
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720

RESULT 3

AD54112
ID AAD54112 standard; DNA; 720 BP.

XX AC
AC AAD54112;

XX DT 17-JUN-2003 (first entry)

XX DE Escherichia coli cytosine nucleoside phosphorylase DNA.

XX KW Cytosine nucleoside phosphorylase; cytosine nucleoside compound;
KW pharmaceutical; enzyme; gene; ds.

XX OS Escherichia coli.

XX FH Key Location/Qualifiers
FT CDS 1..720

XX FT /*tag= a
FT /product= "E. coli cytosine nucleoside phosphorylase"

XX XN EPI254959-A2.

XX XX 06-NOV-2002.

XX XX 01-MAY-2002; 2002EP-00253075.

XX XX 01-MAY-2001; 2001JP-00134352.

XX XX (MITA) MITSUI CHEM INC.

XX PI Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;
PI Ishibashi H, Nagahara K, Fukui Y;

XX XX WPI; 2003-158787/16.

DR P-PSDB; AAE35397.

XX Producing cytosine nucleoside compound by reacting sugar phosphate and
PT cytosine or cytosine derivative in the presence of cytosine nucleoside
PT phosphorylase.

XX PS Disclosure; Page 17-18; 32pp; English.

XX The invention relates to a method for producing cytosine nucleoside
CC compound by reacting sugar phosphate and cytosine or cytosine derivative
CC in the presence of cytosine nucleoside phosphorylase. The method is used
CC to produce cytosine nucleoside compounds which are intermediates for
CC pharmaceuticals. The present sequence is Escherichia coli cytosine
CC nucleoside phosphorylase DNA

XX SQ Sequence 720 BP; 172 A; 190 C; 187 G; 171 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 8; Length 720;

Best Local Similarity 99.7%; Pred. No. 9.6e-217;

Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGTAGTTTGGATGCA 60

Db 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGTAGTTTGGATGCA 60

Qy 61 GCGACCCGCTGCGTGGCAAGTATATTGCTGAAACTTTTCTTGAAGATGCCCGTGAAGTG 120

Db 61 GCGACCCGCTGCGTGGCAAGTATATTGCTGAAACTTTTCTTGAAGATGCCCGTGAAGTG 120

Qy 121 AACAAAGTTCGGGTATGCTGGGCTTACCGGTATTAACAAAGCGCGCAAAATTTCCGTA 180

Db 121 AACAAAGTTCGGGTATGCTGGGCTTACCGGTATTAACAAAGCGCGCAAAATTTCCGTA 180

Qy 181 ATGGGTCAACGCTGTTGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT 240

Db 181 ATGGGTCAACGCTGTTGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT 240

Qy 241 TTGGCGTGAAGAAATTTATCCCGTGGGTCTGTCGGCGAGTTCTGCGCGACGTAAAA 300

Db 241 TTGGCGTGAAGAAATTTATCCCGTGGGTCTGTCGGCGAGTTCTGCGCGACGTAAAA 300

Qy 301 CTGGCGAGCTGTTATCGGTATGGGTGCTGCAACCGATTCAGAAATTAACCGCATCGT 360

Db 301 CTGGCGAGCTGTTATCGGTATGGGTGCTGCAACCGATTCAGAAATTAACCGCATCGT 360

Qy 361 TTTAAAGACCATGACTTTTGGCGTATCGCTGACTTCGACATGGTGGTAAAGCGAGTAGAT 420

Db 361 TTTAAAGACCATGACTTTTGGCGTATCGCTGACTTCGACATGGTGGTAAAGCGAGTAGAT 420

Qy 421 GCAGCTAAAGCACTGGGTATTTGATGCTGCGGTGGGTAAACCTGTTCTCGGCTGACCTGTC 480

Db 421 GCAGCTAAAGCACTGGGTATTTGATGCTGCGGTGGGTAAACCTGTTCTCGGCTGACCTGTC 480

Qy 481 TACTCTCGGACGGCGAAATGTTGACGTGATGAAAAATACGGCATTTCTGGCGTGGAA 540

Db 481 TACTCTCGGACGGCGAAATGTTGACGTGATGAAAAATACGGCATTTCTGGCGTGGAA 540

Qy 541 ATGGAAGCGGTGGTATCTACGGGTGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600

Db 541 ATGGAAGCGGTGGTATCTACGGGTGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATC 600

Qy 601 TGCACCGTATCTGACCACTCCGCACTCACGAGCAGACCACTGCGGCTGAGCGTCAGACT 660

Db 601 TGCACCGTATCTGACCACTCCGCACTCACGAGCAGACCACTGCGGCTGAGCGTCAGACT 660

Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720

Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720

RESULT 4
ADL27841

ID ADL27841 standard; DNA; 720 BP.

XX ADL27841;
AC 20-MAY-2004 (first entry)
DT E coli deoD coding sequence.
XX
DE ds; gene; selection method; yield; pathogen resistance;
KW nutritional quality; stress resistance; plant; transgenic.
XX
XX Escherichia coli.
OS
XX WO2004013333-A2.
XX 12-FEB-2004.
XX 18-JUL-2003; 2003WO-EP007877.
PF 26-JUL-2002; 2002DE-01034287.
XX (BADI) BASF PLANT SCI GMBH.
XX Kock M, Frank M, Badur R;
XX WPI; 2004-157134/15.
DR P-PSDB; ADL27842.
XX
XX Method for transforming plant cells, useful for preparing transgenic
PT plants for e.g. food or production of chemicals, with selection based on
PT suppressing toxic effects of a marker protein.
XX
XX Disclosure; Page 139-140; 201pp; German.
XX
XX The present invention relates to a method for preparing transformed plant
CC cells or organisms, which comprises first transfecting a population of
CC plant cells, already containing at least one marker protein that can
CC exert a direct or indirect toxic effect, with at least one nucleic acid
CC sequence together with at least one compound able to reduce expression,
CC amount, activity and/or function of the marker protein. Transfected cells
CC that contain the nucleic acid in their genome and, because of the
CC activity of the compound have a growth advantage relative to non-
CC transformed cells are selected, under conditions where the marker protein
CC exerts its toxic effect on non-transformed cells. The method is
CC especially used to produce transgenic plants, especially those having an
CC advantageous phenotype, e.g. better nutritional quality for humans or
CC animals, production of selected chemicals or pharmaceuticals, increased
CC resistance to pathogens or environmental stress or higher yields. The
CC present sequence is a coding sequence shown in the exemplification of the
XX invention.
XX
SQ Sequence 720 BP; 172 A; 190 C; 187 G; 171 T; 0 U; 0 Other;
Query Match 99.6%; Score 716.8; DB 12; Length 720;
Best Local Similarity 99.7%; Pred. No. 9.6e-217;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACATTATGCGAAGTATATGCTGAAACTTTCCCTTGAAGATGCCCGTGAAGTG 60
DB 1 ATGGCTACCCACACATTATGCGAAGTATATGCTGAAACTTTCCCTTGAAGATGCCCGTGAAGTG 60
QY 61 GGCACCCGCTGCGTGGGATGATATGCTGAAACTTTCCCTTGAAGATGCCCGTGAAGTG 120
DB 61 GGCACCCGCTGCGTGGGATGATATGCTGAAACTTTCCCTTGAAGATGCCCGTGAAGTG 120
QY 121 AACACGTTGCGGTATGCTGGGCTTACCGGTACTTACAAAGCCCGCAAAATTTCCGTA 180
DB 121 AACACGTTGCGGTATGCTGGGCTTACCGGTACTTACAAAGCCCGCAAAATTTCCGTA 180
QY 181 ATGGGTACCGGTGTTGGTATCCGCTCTGCTTACACCAAGCAACTGATCACCAGAT 240
DB 181 ATGGGTACCGGTATGGGTATCCCGTCTGCTTACACCAAGCAACTGATCACCAGAT 240
QY 241 TTCGGCGTGAAGAAATTTATCCGCGTGGGTTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300

Db 241 TTCCGGGTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGACGTAAAA 300
QY 301 CTGGCGAGTCTGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTTAAACCGCATCCGT 360
Db 301 CTGGCGAGTCTGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTTAAACCGCATCCGT 360
QY 361 TTTAAAGACCATGACATTTTCCGCTATCGCTGACTTCGACATGCTGCGTAAACCGCATAGAT 420
Db 361 TTTAAAGACCATGACATTTTCCGCTATCGCTGACTTCGACATGCTGCGTAAACCGCATAGAT 420
QY 421 GCAGCTAAAGCACTGGGTATGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
Db 421 GCAGCTAAAGCACTGGGTATGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
QY 481 TACTCTCCGACCGCGAATTTTCGACGTGATGGAATAATACGCGCATTTCTCGCGCTGGAA 540
Db 481 TACTCTCCGACCGCGAATTTTCGACGTGATGGAATAATACGCGCATTTCTCGCGCTGGAA 540
QY 541 ATGGAAGCGCTGCTGATCTACGCGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 600
Db 541 ATGGAAGCGCTGCTGATCTACGCGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 600
QY 601 TGCACCGTATCTGACCATTCGCACTCAGGACAGACCACTCCCGCTGAGCGTCAGACT 660
Db 601 TGCACCGTATCTGACCATTCGCACTCAGGACAGACCACTCCCGCTGAGCGTCAGACT 660
QY 661 ACCTTCAAGACATGATCAAAATCGCACCTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 661 ACCTTCAAGACATGATCAAAATCGCACCTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
RESULT 5
ABL51595
ID ABL51595 standard; DNA; 752 BP.
XX
AC ABL51595;
XX
DT 04-JUL-2002 (first entry)
XX
DE Purine nucleoside phosphorylase related DNA sequence.
XX
XX Purine nucleoside phosphorylase; enzyme; molecular biology;
KW biotechnology; genetic engineering; gene; ds.
XX
OS Escherichia coli.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT primer_bind 7..35
FT /*tag= a
FT /note= "oligonucleotide binding site"
FT primer_bind complement(709..742)
FT /*tag= b
FT /note= "oligonucleotide binding site"
XX
XX RU2179188-C2.
XX
XX 10-FEB-2002.
XX
XX 03-MAR-2000; 2000RU-00105214.
XX
XX 03-MAR-2000; 2000RU-00105214.
XX
XX (ASBI=) AS RUSSIA BIO-ORGANIC CHEM INST.
XX
XX Bspivov RS, Gurevich AI, Miroshnikov AI, Chuvikovskii DV;
XX WPI; 2002-265551/31.
XX
XX Production of recombinant purine nucleoside phosphorylase, recombinant
PT plasmid DNA pERPUPH01 and strain Escherichia coli BL21(DE3) perpuh01 for
PT its realization.

XX Disclosure; Col 9-10; Opp; Russian.

CC The present invention describes a recombinant plasmid DNA pERUPH01 which encodes an amino acid sequence of an Escherichia coli purine nucleoside phosphorylase. The recombinant plasmid DNA pERUPH01 consists of: a NcoI/EcoRI-fragment of plasmid pET23d DNA containing a promoter and a transcription terminator of T7 RNA polymerase, a translation enhancer of gene 10 of phage T7, a beta-lactamase gene and a NcoI/EcoRI-fragment of a DNA containing the Escherichia coli purine nucleoside phosphorylase gene sequence adapted to these sites. The recombinant plasmid DNA pERUPH01 can be used in molecular biology, biotechnology and genetic engineering.

CC The present sequence represents a nucleotide sequence given in the exemplification of the present invention

XX

SQ Sequence 752 BP; 184 A; 194 C; 195 G; 179 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 6; Length 752;

Best Local Similarity 99.7%; Pred. No. 9.8e-217;

Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGTAGTTTGTGATGCCA 60

DB |||||||

QY 13 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGTAGTTTGTGATGCCA 72

DB |||||||

QY 61 GCGACCCCGCTGCGTGCAGAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 120

DB |||||||

QY 73 GCGACCCCGCTGCGTGCAGAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 132

DB |||||||

QY 121 AACACGTTCCGGTATGCTGGGCTTCCCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180

DB |||||||

QY 133 AACACGTTCCGGTATGCTGGGCTTCCCGGTACTTACAAAGCGCGCAAAATTTCCGTA 192

DB |||||||

QY 181 ATGGGTACCGGTGTTGGTATCCCGTCTGCTCATCTACACCAAGAACTGATCACCGAT 240

DB |||||||

QY 193 ATGGGTACCGGTATGGGTATCCCGTCTGCTCATCTACCAAGAACTGATCACCGAT 252

DB |||||||

QY 241 TTCGCGGTGAAGAAATATCCCGTGGGTTCTCTGGCGCAGTTCTCCGCGACGTAAAA 300

DB |||||||

QY 253 TTCGCGGTGAAGAAATATCCCGTGGGTTCTCTGGCGCAGTTCTCCGCGACGTAAAA 312

DB |||||||

QY 301 CTGCGCGAGTCGTTATCGGTATGGGTGCTGCACCGATTCGAAATTAACCGCATCCGT 360

DB |||||||

QY 313 CTGCGCGAGTCGTTATCGGTATGGGTGCTGCACCGATTCGAAATTAACCGCATCCGT 372

DB |||||||

QY 361 TTTAAGACCATGACTTTGGCGCTATCGTGCATCTCGACATGTCGTAACGCATGAT 420

DB |||||||

QY 373 TTTAAGACCATGACTTTGGCGCTATCGTGCATCTCGACATGTCGTAACGCATGAT 432

DB |||||||

QY 421 GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTT 480

DB |||||||

QY 433 GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTT 492

DB |||||||

QY 481 TACTCTCGGACGGCGAAATGTTTCAGCTGATGGAATAATACGGCATTTCTCGCGGTGAA 540

DB |||||||

QY 493 TACTCTCGGACGGCGAAATGTTTCAGCTGATGGAATAATACGGCATTTCTCGCGGTGAA 552

DB |||||||

QY 541 ATGGAAGCGGCTGCTGATCTACGGGCTGCTGCAGAAATTTGGCGGAAGCCCTGACCATC 600

DB |||||||

QY 553 ATGGAAGCGGCTGCTGATCTACGGGCTGCTGCAGAAATTTGGCGGAAGCCCTGACCATC 612

DB |||||||

QY 601 TGCACCGTATCTGACCATCGCATCTCAGCAGCAGACCACTCGCGCTGAGCGTCACTACT 660

DB |||||||

QY 613 TGCACCGTATCTGACCATCGCATCTCAGCAGCAGACCACTCGCGCTGAGCGTCACTACT 672

DB |||||||

QY 661 ACCTTCAACGACATGATCAAAATCGCATTCGATTCGTTCTGCTGGCGGATTAAGAGTAA 720

DB |||||||

QY 673 ACCTTCAACGACATGATCAAAATCGCATTCGATTCGTTCTGCTGGCGGATTAAGAGTAA 732

DB |||||||

RESULT 6

AAA51632

ID AAA51632 standard; DNA; 3383 BP.

XX AAA51632;

XX AC

DT 31-OCT-2000 (first entry)

DE Plasmid pGM678 containing E. coli deoD gene.

XX Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD; Genetically modified organism; catalyst; transglycosylation; nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis; anti-viral; anti-tumour; mesophilic bacterium; ss.

XX Escherichia coli.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT misc_RNA 1..230

FT /tag= a

FT /label= pUC18_sequence

FT CDS 216..952

FT /tag= b

FT /product= "lacZ-deoD_fused_gene"

FT misc_RNA 961..3383

FT /tag= c

FT /label= pUC18_sequence

XX WO200039307-A2.

XX

XX 06-JUL-2000.

XX

XX 23-DEC-1999; 99WO-EP010416.

XX

XX 23-DEC-1998; 98IT-MI002792.

XX (NORP-) NORPHARMA SPA.

XX

PI Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;

XX WPI; 2000-452402/39.

XX

PT Recombinant expression vectors used to transform cells for the production of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside phosphorylase.

PT

PS Claim 14; Page 47-48; 72pp; English.

XX

CC Host cells genetically modified to express uridine phosphorylase (Udp) and purine nucleoside phosphorylase (PNP) or their corresponding crude or purified extracts, either separately or in combination are used as catalysts of transglycosylation reactions between a donor nucleoside and an acceptor base, for preparing nucleoside analogues containing heterocyclic systems with purine and/or pyrimidine bases substituted by one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate sugars by phosphorolysis reactions and for producing nucleosides and modified nucleoside analogues (all claimed). The modified or natural nucleosides are used directly or as intermediates in the preparation of drugs with anti-viral and anti-tumour activity and for preparing oligonucleotides for therapeutic or diagnostic use. The recombinant host cells are produced using recombinant plasmid expression vectors comprising at least one gene sequence of a mesophilic bacterium coding for a polypeptide with UDP activity and/or enzyme PNP activity and at least one gene sequence coding for antibiotic resistance. In particular, the E. coli udp and deoD genes are used. Recombinant strains produced using the vectors express polypeptides with enzyme Udp and PNP activity in large amounts, e.g. 340-1040 times higher Udp activity and 120-200 times higher PNP activity than non-transformed corresponding wild type strains

XX

SQ Sequence 3383 BP; 826 A; 868 C; 855 G; 834 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 3; Length 3383;

Best Local Similarity 99.7%; Pred. No. 2e-216;

Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTGATGCCA	60
Db	240	ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTGATGCCA	299
Qy	61	GGCGACCCGCTGCGTGGGAAGTATATTGCTGAACCTTTCCCTTGAAGATGCCCGTGAAGTG	120
Db	300	GGCGACCCGCTGCGTGGGAAGTATATTGCTGAACCTTTCCCTTGAAGATGCCCGTGAAGTG	359
Qy	121	AACAACGTTCCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGTA	180
Db	360	AACAACGTTCCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGTA	419
Qy	181	ATGGGTACCGGTGTTGGTATCCCGTCTGCTGCATCTACACCAAGAACTGATCACCGAT	240
Db	420	ATGGGTACCGGTATGGGTATCCCGTCTGCTGCATCTACACCAAGAACTGATCACCGAT	479
Qy	241	TTCCGCGTGAAGAAAAATTTATCCGCGTGGGTTTCCTGTGGCGAGTTCTGCCGACGTAAAA	300
Db	480	TTCCGCGTGAAGAAAAATTTATCCGCGTGGGTTTCCTGTGGCGAGTTCTGCCGACGTAAAA	539
Qy	301	CTGGCGACGTCGTATCGGTATGGGTGCTGCGACCGATTCCAAAGTTAACCGCATCCGT	360
Db	540	CTGGCGACGTCGTATCGGTATGGGTGCTGCGACCGATTCCAAAGTTAACCGCATCCGT	599
Qy	361	TTTAAAGACCATGACTTTGCCGCTATCGTGACTTCGACATGSGTGCCTAACGCAGTAGAT	420
Db	600	TTTAAAGACCATGACTTTGCCGCTATCGTGACTTCGACATGSGTGCCTAACGCAGTAGAT	659
Qy	421	GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTT	480
Db	660	GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTT	719
Qy	481	TACTCTCGGACGGGAAATGTTTCGACGTGATGGAATAATACGGCATTCGCGGTGGAA	540
Db	720	TACTCTCGGACGGGAAATGTTTCGACGTGATGGAATAATACGGCATTCGCGGTGGAA	779
Qy	541	ATGGAAGCGGCTGTATCTACGGGCTGCTGTCAGAAATTTGGCGGAAGCCCTGACCATC	600
Db	780	ATGGAAGCGGCTGTATCTACGGGCTGCTGTCAGAAATTTGGCGGAAGCCCTGACCATC	839
Qy	601	TGCACCGTATCTGACCATCCGACACTCAGCAGCAGACCACTGCGGCTGAGCGTCAGACT	660
Db	840	TGCACCGTATCTGACCATCCGACACTCAGCAGCAGACCACTGCGGCTGAGCGTCAGACT	899
Qy	661	ACCTTCAACGACATGATCAAAATCGCATGGAATCCGTTCTGCTGGGCGATAAAGAGTAA	720
Db	900	ACCTTCAACGACATGATCAAAATCGCATGGAATCCGTTCTGCTGGGCGATAAAGAGTAA	959
RESULT 7			
AAAS1634			
ID	AAAS1634 standard; DNA; 4189 BP.		
XX			
AC	AAAS1634;		
XX			
DT	31-OCT-2000 (first entry)		
XX			
DE	Plasmid pGM712 containing udp and deoD genes.		
XX			
KW	Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;		
KW	Genetically modified organism; catalyst; transglycosylation; anti-viral;		
KW	nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;		
XX	anti-tumour; mesophilic bacterium; tetracycline resistance; ss.		
OS	Escherichia coli.		
OS	Synthetic.		
XX	Chimeric.		
Key	Location/Qualifiers		
FT	1..242		
FT	/*tag= a		
FT	/label= pUC18_sequence		

FT	misc_RNA	243..1021	
FT		/*tag= b	
FT		/label= UDP_gene_sequence	
FT	misc_RNA	1022..1025	
FT		/*tag= c	
FT		/label= pUC18_sequence	
FT	misc_RNA	1026..1036	
FT		/*tag= d	
FT		/label= pBAD24_sequence	
FT	misc_RNA	1037..1766	
FT		/*tag= e	
FT		/label= deoD_gene_sequence	
FT	misc_RNA	1767..1792	
FT		/*tag= f	
FT		/label= pBAD24_sequence	
FT	misc_RNA	1793..4189	
FT		/*tag= g	
FT		/label= pUC18_sequence	
XX			
PN	WO200039307-A2.		
XX			
PD	06-JUL-2000.		
XX			
PF	23-DEC-1999; 99WO-EP010416.		
XX			
PR	23-DEC-1998; 98IT-MI002792.		
XX			
PA	(NORP-) NORPHARMA SPA.		
XX			
PI	Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;		
XX			
DR	WPI; 2000-452402/39.		
XX			
PT	Recombinant expression vectors used to transform cells for the production		
PT	of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside		
PT	phosphorylase.		
XX			
PS	Claim 14; Page 51-52; 72pp; English.		
XX			
CC	Host cells genetically modified to express uridine phosphorylase (Udp)		
CC	and purine nucleoside phosphorylase (PNP) or their corresponding crude or		
CC	purified extracts, either separately or in combination are used as		
CC	catalysts of transglycosylation reactions between a donor nucleoside and		
CC	an acceptor base, for preparing nucleoside analogues containing		
CC	heterocyclic systems with purine and/or pyrimidine bases substituted by		
CC	one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate		
CC	sugars by phosphorylase reactions and for producing nucleosides and		
CC	modified nucleoside analogues (all claimed). The modified or natural		
CC	nucleosides are used directly or as intermediates in the preparation of		
CC	drugs with anti-viral and anti-tumour activity and for preparing		
CC	oligonucleotides for therapeutic or diagnostic use. The recombinant host		
CC	cells are produced using recombinant plasmid expression vectors		
CC	comprising at least one gene sequence of a mesophilic bacterium coding		
CC	for a polypeptide with UDP activity and/or enzyme PNP activity and at		
CC	least one gene sequence coding for antibiotic resistance. In particular,		
CC	the E. coli udp and deoD genes are used. Recombinant strains produced		
CC	using the vectors express polypeptides with enzyme UDP and PNP activity		
CC	in large amounts, e.g. 340-1040 times higher UDP activity and 120-200		
CC	times higher PNP activity than non-transformed corresponding wild type		
CC	strains		
XX			
SQ	Sequence 4189 BP; 998 A; 1077 C; 1081 G; 1033 T; 0 U; 0 Other;		
	Query Match 99.6%; Score 716.8; DB 3; Length 4189;		
	Best Local Similarity 99.7%; Pred. No. 2.3e-216;		
	Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
Qy	1	ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTCCCTGACGTAGTTTGTGATGCCA	60
Db	1046	ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTCCCTGACGTAGTTTGTGATGCCA	1105
Qy	61	GGCGACCCGCTGCGTGGGAAGTATATTGCTGAACCTTTCCCTGAAGATGCCCGTGAAGTG	120


```

Db      1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 1132

RESULT 9
AAD59423
ID      AAD59423 standard; DNA; 5013 BP.
XX
AC      AAD59423;
XX
DT      18-DEC-2003 (first entry)
XX
DE      E.coli purine analogue nucleoside phosphorylase DNA.
XX
KW      Cytostatic; virucide; purine cleavage enzyme; tumour; therapy; PNP;
KW      drug screening; purine analogue nucleoside phosphorylase; ds.
XX
OS      Escherichia coli.
XX
PN      US2003077268-A1.
XX
PD      24-APR-2003.
XX
PF      18-JUL-2002; 2002US-00198034.
XX
PR      14-SEP-1993; 93US-00122321.
PR      23-AUG-1996; 96US-00702181.
PR      24-JUN-1997; 97US-00881772.
PR      31-OCT-1997; 97US-0084676P.
PR      30-OCT-1998; 98US-00183188.
XX
(SORS/) SORSCHER E J.
PA      (PARK/) PARKER W B.
PA      (WAUD/) WAUD W.
PA      (BENN/) BENNETT L L.
XX
PI      Sorscher EJ, Parker WB, Waud W, Gadi VK, Bennett LL;
XX
DR      WPI; 2003-755064/71.
XX
XX
PT      Killing (non-)replicating targeted mammalian cells and bystander cells
PT      useful for treating tumor cells and/or virally infected cells comprises
PT      delivering purine nucleotide phosphorylase or hydrolase cleavage enzyme
PT      to cells.
XX
PS      Disclosure; Page 20-23; 40pp; English.
XX
CC      The present invention relates to a method of killing replicating or non-
CC      replicating, targeted mammalian cells and bystander cells, comprising
CC      delivering a purine cleavage enzyme to the targeted mammalian cells and
CC      contacting the targeted cells with a purine cleavage enzyme substrate to
CC      kill the targeted cells when cleaved by the enzyme. The purine cleavage
CC      enzyme is used for killing replicating or non-replicating targeted
CC      mammalian cells and bystander cells. They are used for treating tumour
CC      cells and virally infected cells. The method may also be useful in drug
CC      screening. This sequence represents Escherichia coli PNP (purine analogue
CC      nucleoside phosphorylase) DNA
XX
SQ      Sequence 5013 BP; 1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;

Query Match          99.6%; Score 716.8; DB 10; Length 5013;
Beet Local Similarity 99.7%; Pred. No. 2.5e-216;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 ATGGCTACCCACACATTAATGCGAATGCGGCAATTTCGCTGACGTAGTTTTCATGCCA 60
Db      413 ATGGCTACCCACACATTAATGCGAATGCGGCAATTTCGCTGACGTAGTTTTCATGCCA 472
Qy      61 GCGGACCCGCTGCGTGGCAATATATGCTCAAACTTTCCTTGAAGATGCCGCGAAGTG 120
Db      473 GCGGACCCGCTGCGTGGCAATATATGCTCAAACTTTCCTTGAAGATGCCGCGAAGTG 532
Qy      121 AACACAGTTTCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180

```

```

Db      533 AACACAGTTTCGCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 592
Qy      181 ATGGGTCACGGTGTTGGTATCCCGTCTCTGCTCATCTACACCAAGAACTGATCACCGAT 240
Db      593 ATGGGTCACGGTATGGGTATCCCGTCTCTGCTCATCTACACCAAGAACTGATCACCGAT 652
Qy      241 TTCGGGTGAAGAAAATTTATCCCGCTGGGTTCCTGTGGGCGCAGTTCTGCGCGCACGTAAAA 300
Db      653 TTCGGGTGAAGAAAATTTATCCCGCTGGGTTCCTGTGGGCGCAGTTCTGCGCGCACGTAAAA 712
Qy      301 CTGCGGACGTCGTTATCGGTATGGGTGCTGCACCGATTCACCAAGTTTCCAAAGTTAACCCATCCGT 360
Db      713 CTGCGGACGTCGTTATCGGTATGGGTGCTGCACCGATTCACCAAGTTTAAACCGCATCCGT 772
Qy      361 TTTAAAGACCATGACATTTGCCGCTATCGCTGATGCTGACATGCTGGTAAACGCAAGTAGAT 420
Db      773 TTTAAAGACCATGACATTTGCCGCTATCGCTGATGCTGACATGCTGGTAAACGCAAGTAGAT 832
Qy      421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTGTTTC 480
Db      833 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTGTTTC 892
Qy      481 TACTCTCCGACCGCGGAAATGTTTCGACGTGATCGAATAAATAACGGCATTTTCGCGGTGGAA 540
Db      893 TACTCTCCGACCGCGGAAATGTTTCGACGTGATCGAATAAATAACGGCATTTTCGCGGTGGAA 952
Qy      541 ATGGAAGCGGTGGTATCTACGGCGTCTGTCAGAAATTTGGCGGGAAGCCCTGACCATC 600
Db      953 ATGGAAGCGGTGGTATCTACGGCGTCTGTCAGAAATTTGGCGGGAAGCCCTGACCATC 1012
Qy      601 TGCAACGCTATCTGACCACTCCGCACTCACGAGCAGACCACTCCGCTGAGCGCTCAGACT 660
Db      1013 TGCAACGCTATCTGACCACTCCGCACTCACGAGCAGACCACTCCGCTGAGCGCTCAGACT 1072
Qy      661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db      1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 1132

RESULT 10
ADE86127
ID      ADE86127 standard; DNA; 5013 BP.
XX
AC      ADE86127;
XX
DT      29-JAN-2004 (first entry)
XX
DE      E. coli purine nucleoside phosphorylase (PNP) gene.
XX
KW      Tumour growth inhibition; purine analogue; stable transformation;
KW      prokaryotic host cell; purine nucleoside phosphorylase; PNP;
KW      purine nucleoside hydrolase; tumour cell; virally infected cell;
KW      toxin gene therapy; cytostatic; virucide; gene; ds.
XX
OS      Escherichia coli.
XX
PN      US2003134819-A1.
XX
PD      17-JUL-2003.
XX
PF      18-JUL-2002; 2002US-00198228.
XX
PR      14-SEP-1993; 93US-00122321.
PR      23-AUG-1996; 96US-00702181.
PR      24-JUN-1997; 97US-00881772.
PR      31-OCT-1997; 97US-0084676P.
PR      30-OCT-1998; 98US-00183188.
XX
(SORS/) SORSCHER E J.
PA      (PARK/) PARKER W B.
PA      (WAUD/) WAUD W.
PA      (GADI/) GADI V K.

```


Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGATANAAGTAA 1132

RESULT 11

AAAS1633

ID AAAS1633 standard; DNA; 5495 BP.

XX AC AC

XX AAAS1633;

XX 31-OCT-2000 (first entry)

XX

DE Plasmid pGM707 containing deoD and tetracycline resistance genes.

XX

KW Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;

KW Genetically modified organism; catalyst; transglycosylation; anti-viral;

KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;

KW anti-tumour; mesophilic bacterium; tetracycline resistance; ss.

XX

OS Escherichia coli.

OS Synthetic.

OS Chimeric.

XX

XX Key Location/Qualifiers

FT misc_RNA 1. .230

FT /*tag= a

FT /label= pUC18_sequence

FT 216. .952

FT CDS

FT /*tag= b

FT /product= "lacZ-udp_fused_gene"

FT 961. .978

FT misc_RNA

FT /*tag= c

FT /label= pUC18_sequence

FT 979. .1422

FT misc_RNA

FT /*tag= d

FT /label= pHP45_sequence

FT 1423. .2822

FT misc_RNA

FT /*tag= e

FT /label= pBR322_Tet_gene_sequence

FT 2823. .3090

FT misc_RNA

FT /*tag= f

FT /label= pHP45_sequence

FT 3091. .5495

FT misc_RNA

FT /*tag= g

FT /label= pUC18_sequence

XX

XX WO200039307-A2.

XX

XX 06-JUL-2000.

XX

XX 23-DEC-1999; 99WO-EP010416.

XX

XX 23-DEC-1998; 98IT-MI002792.

PR

XX (NORP-) NORPHARMA SPA.

XX

XX Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;

XX

XX WFI; 2000-452402/39.

DR

XX Recombinant expression vectors used to transform cells for the production

PT of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside

PT phosphorylase.

XX

PS Claim 14; Page 48-50; 72pp; English.

XX

XX Host cells genetically modified to express uridine phosphorylase (Udp)

CC and purine nucleoside phosphorylase (PNP) or their corresponding crude or

CC purified extracts, either separately or in combination are used as

CC catalysts of transglycosylation reactions between a donor nucleoside and

CC an acceptor base, for preparing nucleoside analogues containing

CC heterocyclic systems with purine and/or pyrimidine bases substituted by

CC one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate

CC sugars by phosphorylation reactions and for producing nucleosides and
CC modified nucleoside analogues (all claimed). The modified or natural
CC nucleosides are used directly or as intermediates in the preparation of
CC drugs with anti-viral and anti-tumour activity and for preparing
CC oligonucleotides for therapeutic or diagnostic use. The recombinant host
CC cells are produced using recombinant plasmid expression vectors
CC comprising at least one gene sequence of a mesophilic bacterium coding
CC for a polypeptide with UDP activity and/or enzyme PNP activity and at
CC least one gene sequence coding for antibiotic resistance. In particular,
CC the E. coli udp and deoD genes are used. Recombinant strains produced
CC using the vectors express polypeptides with enzyme UDP and PNP activity
CC in large amounts, e.g. 340-1040 times higher UDP activity and 120-200
CC times higher PNP activity than non-transformed corresponding wild type
CC strains
XX
SQ Sequence 5495 BP; 1335 A; 1416 C; 1460 G; 1284 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 3; Length 5495;
Best Local Similarity 99.7%; Pred. No. 2.6e-216;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCGAATATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 60
Db |||||
Qy 240 ATGGCTACCCACACATTAATGCGAATATGGCGGATTTTCGCTGACGTAGTTTTCATGCCA 299
Db |||||

Qy 61 GCGGACCCGCTGCGTGGCACTATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 120
Db |||||

Qy 300 GCGGACCCGCTGCGTGGCACTATATGCTGAACTTTCTTGAAGATGCCCGTGAAGTG 359
Db |||||

Qy 121 AACACGTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 180
Db |||||

Qy 360 AACACGTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTTCCGTA 419
Db |||||

Qy 181 ATGGGTACCGGTATGGGTATCCGCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 240
Db |||||

Qy 420 ATGGGTACCGGTATGGGTATCCGCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 479
Db |||||

Qy 241 TTCGGCGTGAAGAAATTTATCCGGTGGTTCCTGTGGCGAGTTCGTGGCGAGTAA 300
Db |||||

Qy 480 TTCGGCGTGAAGAAATTTATCCGGTGGTTCCTGTGGCGAGTTCGTGGCGAGTAA 539
Db |||||

Qy 301 CTGGCGGACGTCGTTATCGGTATGGGTGCTGACCGATTCCTCAAGTTTCAACCGATCCGT 360
Db |||||

Qy 540 CTGGCGGACGTCGTTATCGGTATGGGTGCTGACCGATTCCTCAAGTTTCAACCGATCCGT 599
Db |||||

Qy 361 TTTAAAGACATGACTTTGCGGCTATGCTGACATTCGACATGGTGGTAACGAGTAGAT 420
Db |||||

Qy 600 TTTAAAGACATGACTTTGCGGCTATGCTGACATTCGACATGGTGGTAACGAGTAGAT 659
Db |||||

Qy 421 GCAGTAAAGCACGTGGGTATGATGCTGCGTGGGTAACTGTTCTCGGCTGACCTGTC 480
Db |||||

Qy 660 GCAGTAAAGCACGTGGGTATGATGCTGCGTGGGTAACTGTTCTCGGCTGACCTGTC 719
Db |||||

Qy 481 TACTCTCGGACGGCGAAATTTTCGACGTATGGAATAACGGCATTTCTCGGCGTGGAA 540
Db |||||

Qy 720 TACTCTCGGACGGCGAAATTTTCGACGTATGGAATAACGGCATTTCTCGGCGTGGAA 779
Db |||||

Qy 541 ATGGAAGCGGTGGTATCTACGGGTGCTGACAGATTTTGGCGGAAGCCCTGACCATC 600
Db |||||

Qy 780 ATGGAAGCGGTGGTATCTACGGGTGCTGACAGATTTTGGCGGAAGCCCTGACCATC 839
Db |||||

Qy 601 TGCACCGGTATCTGACCATCCGACTCAGCAGCAGACCTGCGGCTGAGCGTCACT 660
Db |||||

Qy 840 TGCACCGGTATCTGACCATCCGACTCAGCAGCAGACCTGCGGCTGAGCGTCACT 899
Db |||||

Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATTAAGAGTAA 720
Db |||||

Qy 900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATTAAGAGTAA 959
Db |||||

RESULT 12

AAA51635

ID AAA51635 standard; DNA; 6301 BP.

XX
AC AAA51635;
XX 31-OCT-2000 (first entry)
XX
DE Plasmid pGM716 containing udp, deoD and tetracycline resistance genes.
XX
KW Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;
KW Genetically modified organism; catalyst; transglycosylation; anti-viral;
KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorylase;
KW anti-tumour; mesophilic bacterium; tetracycline resistance; ss.
XX
OS Escherichia coli.
OS Synthetic.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT misc_RNA 243..1021
FT FT /*tag= a
FT FT /label= udp_gene
FT misc_RNA 1037..1766
FT FT /*tag= b
FT FT /label= deoD_gene
FT misc_RNA 2229..3628
FT FT /*tag= c
FT FT /label= pBR322_Tet_gene_sequence
XX
XX WO200039307-A2.
XX
XX 06-JUL-2000.
XX
XX 23-DEC-1999; 99WO-EP010416.
XX
XX 23-DEC-1998; 98IT-MI002792.
XX
XX (NORP-) NORPHARMA SPA.
XX
XX Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;
XX WPI; 2000-452402/39.
XX
XX Recombinant expression vectors used to transform cells for the production
XX of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside
XX phosphorylase.
XX
XX Claim 14; Page 53-55; 72pp; English.
XX
XX Host cells genetically modified to express uridine phosphorylase (Udp)
XX and purine nucleoside phosphorylase (PNP) or their corresponding crude or
XX purified extracts, either separately or in combination are used as
XX catalysts of transglycosylation reactions between a donor nucleoside and
XX an acceptor base, for preparing nucleoside analogues containing
XX heterocyclic systems with purine and/or pyrimidine bases substituted by
XX one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate
XX sugars by phosphorylation reactions and for producing nucleosides and
XX modified nucleoside analogues (all claimed). The modified or natural
XX nucleosides are used directly or as intermediates in the preparation of
XX drugs with anti-viral and anti-tumour activity and for preparing
XX oligonucleotides for therapeutic or diagnostic use. The recombinant host
XX cells are produced using recombinant plasmid expression vectors
XX comprising at least one gene sequence of a mesophilic bacterium coding
XX for a polypeptide with UDP activity and/or enzyme PNP activity and at
XX least one gene sequence coding for antibiotic resistance. In particular,
XX the E. coli udp and deoD genes are used. Recombinant strains produced
XX using the vectors express polypeptides with enzyme UDP and PNP activity
XX in large amounts, e.g. 340-1040 times higher UDP activity and 120-200
XX times higher PNP activity than non-transformed corresponding wild type
XX strains
XX
SQ Sequence 6301 BP; 1507 A; 1625 C; 1686 G; 1483 T; 0 U; 0 Other;

Query Match 99.6%; Score 716.8; DB 3; Length 6301;
Best Local Similarity 99.7%; Pred. No. 2.8e-216;

		Matches	718;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
QY	1	ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGACGTAGTTTGGTGCCA	60								
Db	1046	ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGACGTAGTTTGGTGCCA	1105								
QY	61	GGCGACCGCTCGCTGCGAAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG	120								
Db	1106	GGCGACCGCTCGCTGCGAAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG	1165								
QY	121	AACAAAGCTTGGCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGAAATTTCCGTA	180								
Db	1166	AACAAAGCTTGGCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGAAATTTCCGTA	1225								
QY	181	ATGGGTACCGGTGTTGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT	240								
Db	1236	ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCGAT	1285								
QY	241	TTGGCGGTGAAGAAATTAATCCCGTGGGTTCTGTGGGCGAGTTCTGCGGCACGTAAAA	300								
Db	1286	TTGGCGGTGAAGAAATTAATCCCGTGGGTTCTGTGGGCGAGTTCTGCGGCACGTAAAA	1345								
QY	301	CTGGCGGACGCTCGTTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAAACCGATCCGT	360								
Db	1346	CTGGCGGACGCTCGTTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAAACCGATCCGT	1405								
QY	361	TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTTCGACATGGTGGTAAACGAGTAGAT	420								
Db	1406	TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTTCGACATGGTGGTAAACGAGTAGAT	1465								
QY	421	GCAGCTAAGCACTGGGTATTGATGCTCGCGTGGTAAACCTGTTCTCCGCTGACCTGTC	480								
Db	1466	GCAGCTAAGCACTGGGTATTGATGCTCGCGTGGTAAACCTGTTCTCCGCTGACCTGTC	1525								
QY	481	TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAAATACGGCAATTCGGCGTGAA	540								
Db	1526	TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAAATACGGCAATTCGGCGTGAA	1585								
QY	541	ATGGAAGCGGCTGGTATCTACGGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC	600								
Db	1586	ATGGAAGCGGCTGGTATCTACGGCTCGCTGAGAAATTTGGCGGAAAGCCCTGACCATC	1645								
QY	601	TGCACCGTATCTGACACATCCGCACTCAGAGCAGACACTGCGCTGAGCGTCAGACT	660								
Db	1646	TGCACCGTATCTGACACATCCGCACTCAGAGCAGACACTGCGCTGAGCGTCAGACT	1705								
QY	661	ACCTTCAACGACATCATAAATCCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA	720								
Db	1706	ACCTTCAACGACATCATAAATCCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA	1765								
RESULT 13											
AAD56238											
ID	AAD56238 standard; DNA; 720 BP.										
XX											
AC	AAD56238;										
XX											
DT	07-AUG-2003 (first entry)										
DE	Escherichia coli mutant PNP enzyme (A157V) encoding DNA #SEQ ID 3.										
XX											
KW	Purine nucleoside phosphorylase; mutant; tumour; PNP; enzyme; cancer;										
KW	leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy;										
KW	gene; ds.										
OS	Escherichia coli.										
OS	Synthetic.										
FH	Key										
CD	1..720										
FT	/tag= a										
FT	/product= "Mutant PNP enzyme"										
XX											

PN	WO2003035012-A2.										
XX											
PD	01-MAY-2003.										
XX											
PF	28-OCT-2002; 2002WO-US034626.										
XX											
PR	26-OCT-2001; 2001US-00035300.										
XX											
PA	(UABR-) UAB RES FOUND.										
PA	(SOUR-) SOUTHERN RES INST.										
XX	(CORR) CORNELL RES FOUND INC.										
PI	Elalick SE, Parker WB, Seerist JA, Sorscher EJ;										
XX											
DR	WPI; 2003-421350/39.										
DR	N-PSDB; AAE37218.										
XX											
PT	Novel purified mutant purine cleaving enzyme having different biological										
PT	activity than a wild-type purine cleaving enzyme, useful for treating a										
XX	pathological condition characterized by abnormal cell growth.										
PS	Claim 12; Page 65-66; 67pp; English.										
XX											
CC	The invention relates to mutant purine nucleoside phosphorylase (PNP)										
CC	enzymes and nucleic acid molecules encoding such enzymes. These mutant										
CC	enzymes have greater biological activity than wild-type enzymes. PNP										
CC	enzymes are useful for treating a pathological condition characterised by										
CC	abnormal cell growth such as that occurring in cancer of the skin.										
CC	breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,										
CC	nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,										
CC	thyroid, testicle and uterus as well as other conditions characterised by										
CC	abnormal cell growth such as myeloid leukaemia, glioblastoma and										
CC	lymphoma. Sequences of the invention are useful for treating metastatic										
CC	solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or										
CC	to activate prodrugs in the treatment of cancer. They are useful for										
CC	treating virally infected cells. They are also useful in gene therapy.										
CC	The present sequence is Escherichia coli mutant PNP enzyme (A157V)										
CC	encoding DNA										
SQ	Sequence 720 BP; 172 A; 189 C; 187 G; 172 T; 0 U; 0 Other;										
Query Match 99.3%; Score 715.2; DB 8; Length 720;											
Best Local Similarity 99.6%; Pred. No. 3.1e-216;											
Matches 717; Conservative 0; Mismatches 3; Indels 0; Gaps 0;											
QY	1	ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGACGTAGTTTGGTGCCA	60								
Db	1	ATGGCTACCCACACATTAATGAGAAATGGCGGATTTTCGCTGACGTAGTTTGGTGCCA	60								
QY	61	GGCGACCGCTCGCTGCGAAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG	120								
Db	61	GGCGACCGCTCGCTGCGAAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG	120								
QY	121	AACAAAGCTTGGGTATGCTGGGCTTCCCGGTATCAAAAGCGCGCAAAATTTCCGTA	180								
Db	121	AACAAAGCTTGGGTATGCTGGGCTTCCCGGTATCAAAAGCGCGCAAAATTTCCGTA	180								
QY	181	ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT	240								
Db	181	ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT	240								
QY	241	TTGGCGGTGAAGAAATTAATCCCGTGGGTTCTGTGGCGAGTTCTGCGGCACGTAAAA	300								
Db	241	TTGGCGGTGAAGAAATTAATCCCGTGGGTTCTGTGGCGAGTTCTGCGGCACGTAAAA	300								
QY	301	CTGGCGGACGCTCGGTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAAACCGATCCGT	360								
Db	301	CTGGCGGACGCTCGGTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAAACCGATCCGT	360								
QY	361	TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTTCGACATGGTGGTAAACGAGTAGAT	420								
Db	361	TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTTCGACATGGTGGTAAACGAGTAGAT	420								

QY 421 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
DB |||||
QY 421 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
DB |||||
QY 481 TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAATAACGGCATCTCTCGCGTGGAA 540
DB |||||
QY 541 ATGAAGCGGTGTTATCTACGGGTGCTGCGAATTTGGCGGAAGCCCTGACCATC 600
DB |||||
QY 601 TGCACCGTATCTGACCAATCCGACATCAGAGCAGACCACTGCGCGTGAAGCCCTGACACT 660
DB |||||
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATAAAGAGTAA 720
DB |||||

RESULT 14

AAF55446
ID AAF55446 standard; DNA; 720 BP.

AC AAF55446;

DT 29-MAY-2001 (first entry)

DE Nucleotide sequence of a purine nucleoside phosphorylase.

XX Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;
KW purine nucleoside phosphorylase; phosphopentose mutase;
KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;
KW deoxyribokinase; nucleoside 2-deoxyriboseyltransferase; ss.

XX Escherichia coli.

XX Key Location/Qualifiers

FT CDS 1..720

FT /*tag= a

FT /product= "purine nucleoside phosphorylase"

XX WO200114566-A2.

PD 01-MAR-2001.

XX 18-AUG-2000; 2000WO-EP008088.

XX 20-AUG-1999; 99EP-00116425.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

XX (INSP) INST PASTEUR.

XX (PHAR-) PHARMA-WALDHOFF GMBH & CO KG.

XX Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;

PI Marliere P, Pochet S;

XX WPI; 2001-235026/24.

DR P-PSDB; AAB67590.

XX In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting
PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside
PT and an inorganic phosphate.

XX Disclosure; Page 63-64; 73pp; English.

XX The present sequence encodes a purine nucleoside phosphorylase enzyme.
CC This enzyme is involved in the biosynthesis of deoxyribonucleosides, and
CC is used in the method of the invention. The specification describes a
CC method for the in vitro enzymatic synthesis of deoxyribonucleosides. The
CC method comprises reacting deoxyribose 1-phosphate and a nucleobase to
CC form a deoxyribonucleoside and an inorganic phosphate. Enzymes which may

CC be used in the method of the invention include thymidine phosphorylase,
CC purine nucleoside phosphorylase, phosphopentose mutase, phosphopentose
CC aldolase, fructose 1,6-diphosphate aldolase, deoxyribokinase, and
CC nucleoside 2-deoxyriboseyltransferase

XX SQ Sequence 720 BP; 174 A; 190 C; 185 G; 171 T; 0 U; 0 Other;

Query Match 99.1%; Score 713.6; DB 5; Length 720;

Best Local Similarity 99.4%; Pred. No. 9.9e-216;

Matches 716; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAAATGCAGAAATGGGCGATTTTCGCTACGCTAGTTTGTATGCCA 60

DB 1 ATGGCTACCCACACATTAAATGCAGAAATGGGCGATTTTCGCTACGCTAGTTTGTATGCCA 60

QY 61 GGGCAGCCGCTGCTGCGAAGTATATGCTGAAACTTCTCTGAAGTCCCGTGAAGTG 120

DB 61 GGGCAGCCGCTGCTGCGAAGTATATGCTGAAACTTCTCTGAAGTCCCGTGAAGTG 120

QY 121 AACAAAGTTCGCGGTATGCTGGGCTTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180

DB 121 AACAAAGTTCGCGGTATGCTGGGCTTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180

QY 181 ATGGGTACCGGTTCGCTATCCGTCCTGCTCCATCTACACAAAGAACTGATCACCGAT 240

DB 181 ATGGGTACCGGTTCGCTATCCGTCCTGCTCCATCTACACAAAGAACTGATCACCGAT 240

QY 241 TTGCGGCTGAGAAATATATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCCGACGTAATA 300

DB 241 TTGCGGCTGAGAAATATATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCCGACGTAATA 300

QY 301 CTGCGGACGCTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 360

DB 301 CTGCGGACGCTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 360

QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGCTGGGTAAACGCACTAGAT 420

DB 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGCTGGGTAAACGCACTAGAT 420

QY 421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480

DB 421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480

QY 481 TACTCTCCGGACGGCGAAATGTTTCGACGTGATCGAATAATACCGCATTTCTCGCGCTGGAA 540

DB 481 TACTCTCCGGACGGCGAAATGTTTCGACGTGATCGAATAATACCGCATTTCTCGCGCTGGAA 540

QY 541 ATGGAAGCGGCTGGTATCTACGGCGTGGTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 600

DB 541 ATGGAAGCGGCTGGTATCTACGGCGTGGTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 600

QY 601 TGCACCGTATCTGACCAATCCGACATCCGACCTCAGAGCAGACCACTCCCGCTGAGCGTCACT 660

DB 601 TGCACCGTATCTGACCAATCCGACATCCGACCTCAGAGCAGACCACTCCCGCTGAGCGTCACT 660

QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATAAAGAGTAA 720

DB 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATAAAGAGTAA 720

RESULT 15

AAA51641

ID AAA51641 standard; DNA; 3066 BP.

XX AAA51641;

XX 31-OCT-2000 (first entry)

XX pGM747 containing deob without upstream promoter.

XX Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deob;
KW Genetically modified organism; catalyst; transglycosylation; anti-viral;
KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;

anti-tumour; mesophilic bacterium; tetracycline resistance; ss.

Escherichia coli.

Synthetic.

Chimeric.

WO200039307-A2.

06-JUL-2000.

23-DEC-1999; 99WO-EP010416.

23-DEC-1998; 98IT-MI002792.

(NORP-) NORPHARMA SPA.

Beetetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;

WPI; 2000-452402/39.

Recombinant expression vectors used to transform cells for the production of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside phosphorylase.

Claim 14; Page 66-67; 72pp; English.

Host cells genetically modified to express uridine phosphorylase (Udp) and purine nucleoside phosphorylase (PNP) or their corresponding crude or purified extracts, either separately or in combination are used as catalysts of transglycosylation reactions between a donor nucleoside and an acceptor base, for preparing nucleoside analogues containing heterocyclic systems with purine and/or pyrimidine bases substituted by one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate sugars by phosphorolysis reactions and for producing nucleosides and modified nucleoside analogues (all claimed). The modified or natural nucleosides are used directly or as intermediates in the preparation of drugs with anti-viral and anti-tumour activity and for preparing oligonucleotides for therapeutic or diagnostic use. The recombinant host cells are produced using recombinant plasmid expression vectors comprising at least one gene sequence of a mesophilic bacterium coding for a polypeptide with Udp activity and/or enzyme PNP activity and at least one gene sequence coding for antibiotic resistance. In particular, the E. coli udp and deoB genes are used. Recombinant strains produced using the vectors express polypeptides with enzyme Udp and PNP activity in large amounts, e.g. 340-1040 times higher Udp activity and 120-200 times higher PNP activity than non-transformed corresponding wild type strains

Sequence 3066 BP; 746 A; 767 C; 772 G; 746 T; 0 U; 35 Other;

Query Match 88.9%; Score 639.8; DB 3; Length 3066;

Best Local Similarity 98.8%; Pred. No. 5.6e-192;

Matches 718; Conservative 0; Mismatches 2; Indels 7; Gaps 7;

Qy	1	ATGGCTACCCACACATTAAATGCAGAAATGGCGATTTCGCTACGTTAGTTTTCATGCCA	60
Db	102	ATGGCTACCCACACATTAAATGCAGAAATGGCGATTTCGCTACGTTAGTTTTCATGCCA	161
Qy	61	GGCGACCGCTCGCTGCGAAGTATATGCTGAAATTTCC-TTGAAGATGCCCGTGAAGT	119
Db	162	GGCGACCGCTCGCTGCGAAGTATATGCTGAAATTTCCNTTGAAGATGCCCGTGAAGT	221
Qy	120	GAACAACTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGT	179
Db	222	GAACAACTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGT	281
Qy	180	AATGGGTACCGTGTGGTAT-CCCGTCTCTCTCATCTACACCAAGAACTGATCACCG	238
Db	282	AATGGGTACCGTATGGGTATNCCGTCCTGCTCATCTACACCAAGAACTGATCACCG	341
Qy	239	ATTTCCGGGTGAAGAAATATCCGCTGGGTTCTCTGGCGCAGTTCTGCGCACGTAA	298
Db	342	ATTTCCGGGTGAAGAAATATCCGCTGGGTTCTCTGGCGCAGTTCTGCGCACGTAA	401

Qy	299	AA-CTGCGCAGCGTCGTTATCGGTATCGGTGCTGCAACCGATTCCAAAGTTTAAACGCAATC	357
Db	402	AANCTGCGCAGCGTCGTTATCGGTATCGGTGCTGCAACCGATTCCAAAGTTTAAACGCAATC	461
Qy	358	CGTTTAAAGACCATGACTTTGGCGCTATCGTGACTTCGACA-TGGTGGCTTAACGCAGT	416
Db	462	CGTTTAAAGACCATGACTTTGGCGCTATCGTGACTTCGACA-TGGTGGCTTAACGCAGT	521
Qy	417	AGATGCAGCTAAAGCACTGGGTATTGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCT	476
Db	522	AGATGCAGCTAAAGCACTGGGTATTGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCT	581
Qy	477	GTTCTACTCTCCGGACCGCGAAAT-GTTTCGACGTGATGGAATAATACGGCAATTCGCGCG	535
Db	582	GTTCTACTCTCCGGACCGCGAAATGTTTCGACGTGATGGAATAATACGGCAATTCGCGCG	641
Qy	536	TGGAATGGAGCGGCTGTTATCTACCGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGA	595
Db	642	TGGAATGGAGCGGCTGTTATCTACCGCTCGCTGCAGAAATTTGGCGCGAAAGCCCTGA	701
Qy	596	CCATC-TGCACCGTATCTGACCACTCCGCACTCACGAGCAGACCACTGCCGCTGAGCGT	654
Db	702	CCATC-TGCACCGTATCTGACCACTCCGCACTCACGAGCAGACCACTGCCGCTGAGCGT	761
Qy	655	CAGACTACCTTCAACGACATGATCAAAATCGCACTCGAATCCGTTTC-TGCTGGGCGATAA	713
Db	762	CAGACTACCTTCAACGACATGATCAAAATCGCACTCGAATCCGTTTC-TGCTGGGCGATAA	821
Qy	714	AGAGTAA 720	
Db	822	AGAGTAA 828	

Search completed: July 22, 2005, 19:37:30
Job time : 727 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 22, 2005, 17:50:12 ; Search time 229 Seconds
(without alignments)
5144.625 Million cell updates/sec

Title: US-10-035-300A-1
Perfect score: 720
Sequence: 1 atgctaccacacattaa.....tgctggcgataaagagtaa 720

Scoring table: IDENTITY NUC
Gapop 10,0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	716.8	99.6	5013	US-09-183-188B-5	Sequence 5, Appli
2	558.4	77.6	732	US-09-489-039A-1631	Sequence 1631, Ap
3	420.6	55.9	1830121	US-09-543-681A-2371	Sequence 2371, Ap
C 4	402.6	55.9	1830121	US-09-557-884-1	Sequence 1, Appli
C 5	402.6	55.9	1830121	US-09-643-990A-1	Sequence 1, Appli
6	265	36.8	744	US-09-902-540-3568	Sequence 3568, Ap
C 7	265	36.8	19019	US-09-902-540-1171	Sequence 1171, Ap
C 8	243.4	33.8	640681	US-09-790-988-1	Sequence 1, Appli
9	210.8	29.3	5301	US-08-956-171E-443	Sequence 443, App
10	210.8	29.3	5301	US-08-781-986A-443	Sequence 443, App
11	204.2	28.4	711	US-09-583-110-37	Sequence 37, Appl
12	204.2	28.4	714	US-09-107-433-1967	Sequence 1967, Ap
13	204.2	28.4	13440	US-08-961-527-128	Sequence 128, App
14	198.6	27.6	735	US-09-107-532A-1185	Sequence 1185, Ap
15	180.8	25.1	2962	US-09-710-279-3511	Sequence 3511, Ap
16	179.2	24.9	732	US-09-134-001C-1402	Sequence 1402, Ap
17	174.6	24.2	714	US-09-134-000C-2608	Sequence 2608, Ap
18	157.4	21.9	2006	US-08-956-171E-378	Sequence 378, App
19	157.4	21.9	2006	US-08-781-986A-378	Sequence 378, App
C 20	141.4	19.6	5739	US-09-634-238-1	Sequence 1, Appli
21	103.4	14.8	1124	US-08-809-254A-3	Sequence 3, Appli
22	99.4	13.8	580073	US-08-545-528D-1	Sequence 1, Appli
23	39.8	5.5	483	US-09-252-991A-7028	Sequence 7028, Ap
C 24	39.8	5.5	2016	US-09-252-991A-7070	Sequence 7070, Ap
25	39.8	5.5	6855	US-09-252-991A-6976	Sequence 6976, Ap
26	37.6	5.2	1830121	US-09-557-884-1	Sequence 1, Appli
27	37.6	5.2	1830121	US-09-643-990A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-183-188B-5
; Sequence 5, Application US/09183188B
; Patent No. 6491905
; GENERAL INFORMATION:
; APPLICANT: Sorscher, Eric J.
; APPLICANT: Parker, William B.
; APPLICANT: Waud, William
; APPLICANT: Gadi, Vijaykrishna K.
; TITLE OF INVENTION: RECOMBINANT E. COLI FOR DELIVERY OF PNP TO TUMOR CELLS
; FILE REFERENCE: UAB-12405/22
; CURRENT APPLICATION NUMBER: US/09/183,188B
; CURRENT FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 08/702,181
; PRIOR FILING DATE: 1996-08-23
; PRIOR APPLICATION NUMBER: 08/122,321
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: 60/064,676
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 5013
; LENGTH: 5013
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-183-188B-5

Query Match 99.6%; Score 716.8; DB 4; Length 5013;
Best Local Similarity 99.7%; Pred. No. 1.3e-225;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

C	28	37	5.1	768	4	US-09-252-991A-9256	Sequence 9256, Ap
C	29	37	5.1	1629	4	US-09-252-991A-9291	Sequence 9291, Ap
C	30	37	5.1	1641	4	US-09-252-991A-9280	Sequence 9280, Ap
C	31	35.8	5.0	822	4	US-09-640-211A-523	Sequence 523, App
C	32	35.6	4.9	1242	4	US-09-902-540-7283	Sequence 7283, Ap
C	33	35.6	4.9	1338	4	US-09-328-352-3122	Sequence 3122, Ap
C	34	35.6	4.9	5054	4	US-09-902-540-687	Sequence 687, App
C	35	35.4	4.9	4998	4	US-09-501-171-5	Sequence 5, Appli
C	36	35.4	4.9	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C	37	35.4	4.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C	38	35.2	4.9	1392	4	US-09-489-039A-6550	Sequence 6550, Ap
C	39	34.8	4.8	786	4	US-09-902-540-4729	Sequence 4729, Ap
C	40	34.8	4.8	28762	4	US-09-902-540-1232	Sequence 1232, Ap
C	41	34.6	4.8	711	4	US-09-107-532A-3162	Sequence 3162, Ap
C	42	34.4	4.8	1566	4	US-09-328-352-1297	Sequence 1297, Ap
C	43	34.4	4.8	239527	4	US-09-949-016-15980	Sequence 15980, A
C	44	33.8	4.7	972	4	US-09-902-540-8031	Sequence 8031, Ap
C	45	33.8	4.7	6715	4	US-09-902-540-818	Sequence 818, App

STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
FILING DATE: 23-AUG-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186PIC1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Query Match 55.9%; Score 402.6; DB 4; Length 1830121;
Best Local Similarity 72.7%; Pred. No. 2.4e-120;
Matches 519; Conservative 0; Mismatches 195; Indels 0; Gaps 0;
QY 7 ACCCCACACATTAATGCAGAAATGGCGATTTTCGCTGACGTAGTTTTCATGCCAGGCGAC 66
DB 540876 ACTCCACATTAATTAACCGCGCTGAAAGTGCATTTGCTGATGTTTAAATGCTGGCGAT 540817
QY 67 CGCGTCGGTGCAGATATATGCTGAAACTTTCTTGAAGATGCCCGTGAAGTGAACAAC 126
DB 540816 CCACCTCGTGCAAAATATATTCAGAAACTTTTTCAGAACTGTTTGAAGTGAACGAAT 540757
QY 127 GTTCGGGGTATCGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTATCGGT 186
DB 540756 GTTCGTAATATGCTGGTTTACTGGAACCTTAAAGGTCGTAAAACTTTATTATGGG 540697
QY 187 CAGCGTGTGGTATCCCGTCTGCTCCATCTACACAAAGAACTGATCACCAGATTTTCGGC 246
DB 540696 CAGCGTATGGGNTTCCATCTTGCTCTATTACGCGAAAGATTAATCACTGAATATGGC 540637
QY 247 GTGAAGAAATATACCGGTGGGTTCTGTGGCGAGTTTCGCGCACGTAATAACTCGC 306
DB 540636 GTGAAGAAATATACCGGTAGTTCGTTCTGTGGCACTGTTTCGTATGGACGTGAAGTGGC 540577
QY 307 GACGTCGTTATCGGTATGGGTGCGTCCACCGATTTCAAAGTTAACCGCATCGGTTTAAA 366
DB 540576 GATGTGATATCGGTTTAAAGTGCATGTACTGATTCAAAGTAAACCGTATTTCGTTTCAA 540517
QY 367 GACCATGACTTTTCGCGTATCGCTGACTTTCGACATGTTGCGTAAAGTGAAGTGCAGCT 426
DB 540516 GATAACGACTTTTNCAGCTATGCTGACTTTGATTTGATTTGGCAAGCGCTGTTTCAAGCTGCA 540457
QY 427 AAAGCACTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCGCTGACCTGTTTACTCT 486
DB 540456 AAAGCAAAAGGTTAAAGTCGTTGCTGTAGTAAATTTATTTCTGGCGGATTTTATCTATACA 540397
QY 487 CCGGACGGCGAAATGTTTCAAGCTGATGGAATAATACGGCATTTCTCGGCGTGGAAATGGA 546
DB 540396 CCGGATGTGGAAATGTTTCGATGTAAATGGAATAATACGGCATTTTATAGGTGTGGAAATGGA 540337

Db 481 TACCAACCCGAGGAGCAGCTCAACGCCACCCCTGGCGCGGATGGGTGTCTCTGGCCGTCGAG 540
Qy 541 ATGGAAGCGGTGTGTATCTACGCGCTCGCTGAGAAATTTGGCGCAAAAGCCCTGACCAATC 600
Db 541 ATGGAAGTGGCGGCTCTACGCGCTGGCGGAGTGGGCGCCCGCGCTGGGGCTG 600
Qy 601 TGCACCGTATCTGACCAATCGGCACTCACGAGCAGACACTGCGCGCTGAGCGTCAGACT 660
Db 601 CTCACGGTGTGGACCAATCATCACCGGGAGAGTCTCACGCCGAGGAGCGGAGAGC 660
Qy 661 ACCTTCAACGACATCATCAAAATCGCACTGGAATCCG 697
Db 661 ACGTTGACGAGATGATTGAACCTCGCCCTGGAGCTG 697

RESULT 7

US-09-902-540-1171/c
; Sequence 1171, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1171
; LENGTH: 19019
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(19019)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1171

Query Match 36.8%; Score 265; DB 4; Length 19019;
Best Local Similarity 61.3%; Pred. No. 4.2e-76;
Matches 427; Conservative 0; Mismatches 270; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGAGTAGTATTGATGCCA 60
Db 16556 ATGGGACTCTCATATCTCCGCTGCCCTGGTACTTTCGCTGAGTGTCTCATGCT 16497
Qy 61 GCGACCCGCTGCGTGGCAAGTATATGCTGAACTTTTCTTGAAGATGCCCGTGAAGTG 120
Db 16496 GCGACCCGCTCCGCGCTCGTTACATCTCCGACCGCTTCTTGGAAAGTGGCCGCGAGGTC 16437
Qy 121 AACACGTTCCGGGTATGCTGGCTTACCGGTACTTACAAGGCCGCAAAATTTCCGTA 180
Db 16436 ACCTCCGGTGGCAACATGCTCGGCTTACCGGGACCTTCCGGGGCGCGGCTGTCG 16377
Qy 181 ATGGGTACCGGTGTGGTATCCGCTCCGCTGCTACACCAAGAACTGATCACCGAT 240
Db 16376 ATGGGGACGCGATGGGGTTCCTTCATCTTCATCTACGCAACCGAGCTCATCAGACG 16317
Qy 241 TTCGGCGTGAAGAAATATTCGCGTGGGTTCCTGTGGCGAGTTCTGCGCGACGTAAAA 300
Db 16316 TATGGGTGCGCGTGATCATCCGCTGGCGAGCTGCGGCGCTGAGCACCGACGTGAAG 16257
Qy 301 CTGGCGAGCTGCTGATCGGTATGGGTGCTGCGACCGGATTCAAAGTTAACCGCATCCGT 360
Db 16256 GTCCGGGAGGTTCATGCTCGACCGGGGCGCGACGGAATCCAAAGTGAATCGGATCGG 16197
Qy 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGCGTAACGAGTAGAT 420
Db 16196 CTGATGGGGCATGACTTCGCCCGCGGTGGCGGACTTCACGCTCGCGCGCGCGCATGGAG 16137

Qy 421 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAACTGTTTCTCGCTGACCTGTC 480
Db 16136 GCGCGGAGCGCGCAACAAGCCTGTGCGCGCGCGCGCTTTCACCTCCGACCTTTC 16077
Qy 481 TACTCTCCGAGCGGGAATGTTGACGTGATGGAATAATACGGCATTTCTCGCGCTGGA 540
Db 16076 TACCACCGCAGGAGAGCTCAACGCCACCTTGGCGCGATGGGTGTCTGCGCGCTCGAG 16017
Qy 541 ATGGAAGCGGTGTGTATCTACGCGGTCTGTCAGAAATTTGGCGCAAAAGCCCTGACCAATC 600
Db 16016 ATGAGGTGCGCGCTCTACGCGTGGCGGAGTGGGCGCCCGCGCTGGGGCTG 15957
Qy 601 TGCACCGTATCTGACCAATCGGCACTCACGAGCAGACACTGCGCGCTGAGCGTCAGACT 660
Db 15956 CTCACGGTGTGGACCAATCATCACCGGGAGAGTCTCACGCCGAGGAGCGGAGAGC 15897
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCG 697
Db 15896 ACGTTGACGAGATGATTGAACCTCGCCCTGGAGCTG 15860

RESULT 8

US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRO
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 33.8%; Score 243.4; DB 4; Length 640681;
Best Local Similarity 59.2%; Pred. No. 4.6e-68;
Matches 415; Conservative 0; Mismatches 286; Indels 0; Gaps 0;
Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGCTGAGTAGTATTGATGCCA 60
Db 575671 ATGTCTACTCCACATATTAATAGTAAAGATGATTTTTCAGATATAGTTTAAATGCG 575612
Qy 61 GCGACCCGCTGCGTGGCAAGTATATGCTGAACTTTTCTTGAAGATGCCCGTGAAGTG 120
Db 575611 GGAGACCCGCTGCTGCAAAATATATGCTGAAATATTTAAAGTAATTTTGTCAAGTA 575552
Qy 121 AACACGTTTCGCGTATGCTGGGCTTACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
Db 575551 AATGATCTCGTTAAATGTTAGCTTATACCGGATTTTATAAATAAGATTTTCAATC 575492
Qy 181 ATGGGTCAAGGTGTGGTATCCCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Db 575491 ATGAGTCATGTAAGGAATACCATCAGCTTCTCTATACAGAGAGTTAATTTATGAA 575432
Qy 241 TTCCGCTGAGAAATATTCGCGTGGGTTCCTGTCGCGCAGTTCTGCGCGCAGTAAA 300
Db 575431 TTTAATGTAAAAAATTTATTCGTATAGGAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 575372
Qy 301 CTGGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
Db 575371 TTACGTGATATAGTAATCAGTATGGGAGCTTCTACTGATTTCTAAGTAATAGTAAGA 575312
Qy 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTTCGACATGGTGGCGTAAACGAGTAGAT 420

[illegible]

RESULT 9

```

US-08-956-171E-443
; Sequence 443, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
;           Gil H. Choi
;           Patrick S. Dillon
;           Craig A. Rosen
;           Steven C. Barash
;           Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Human Genome Sciences, Inc.
;   STREET: 9410 Key West Avenue
;   CITY: Rockville
;   STATE: Maryland
;   COUNTRY: USA
;   ZIP: 20850
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;   COMPUTER: HP Vectra 486/33
;   OPERATING SYSTEM: MSDOS version 6.2
;   SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/956,171E
;   FILING DATE: 20-Oct-1997
;   CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/009,861
;   FILING DATE: January 5, 1996
;   APPLICATION NUMBER: 08/781,986
;   FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Mark J. Hyman
;   REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (240) 314-1224
;   TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 443:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 5301 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 443:
US-08-956-171E-443

```

SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781.986A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Benson, Bob
REGISTRATION NUMBER: 30,446
REFERENCE/DOCKET NUMBER: PB248PP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 443:
SEQUENCE CHARACTERISTICS:
LENGTH: 5301 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-781-986A-443

Query Match 29.3%; Score 210.8; DB 4; Length 5301;
Best Local Similarity 56.5%; Pred. No. 1.5e-58;
Matches 392; Conservative 0; Mismatches 302; Indels 0; Gaps 0;
Qy 4 GCTACCCACACATTAATGACAGAAATGGCGGATTCGCTGACGTAGTTTTTCATGCCAGGC 63
Db 4097 GGTACACACATATTCACCAAAATGGAGTAATAATGCTAAACAGATTAATATGCTGGC 4156
Qy 64 GACCCGCTGCGTGGAAAGTATATTCGTGAATCTTTCCCTTGAAGATGCCCGTGAAGTGAAC 123
Db 4157 GATCCGCTACGTCGCAAAATATTCGTGATAATTTTTAGAAAATGTTGAACAATTTTAAAC 4216
Qy 124 AACGTTCCGGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGTAATG 183
Db 4217 GATGACGTAAATGTTGGTGTACACTGGTACATATAAGGTAAGAAGTTTCTGTAAATG 4276
Qy 184 GGTCAACGCTGTGGTATGCCGCTCCGCTCCTACACCAAGAACTGATCACCGATTTTC 243
Db 4277 GGTTCGTATGGTATTCGAAGTATTTGGTATTTACTCATATGATGATATACACTCTTTT 4336
Qy 244 GCGGTGAAGAAATATACCGCGTGGTTCCTGTGGCGAGTTCTGCCGACGTAAAACTG 303
Db 4337 GATGTAGATACAACTCATTCGTATCGGTTCTTGTGGCGCATTTACAAGAAATGTTAACTTA 4396
Qy 304 CGGACGCTGTATCGGTATGGTGGCTGCGACCGCATTCCAAGTTAACCGCATCGGTTT 363
Db 4397 TACGATGTTATTTATGCAACAAGCTGCATCAACTAAATTCAAATTTATGTAGATCAATAC 4456
Qy 364 AAAGACCATGATTTTGGCGCTATCGCTGACITTCGACATGGTGGCTTAAGCGAGTAGATGCA 423
Db 4457 ATTCAGGTCAATTCGCGCCCTATCGTGACITTCGAGTTAGTAACTAAAGCTAAATAATGTC 4516
Qy 424 GCTAAAGCACTGGGTATTTGATGCTCGGTGGGTAACTGTTTCTCGCTGACCTGTTCTAC 483
Db 4517 GCTGACCAATTCGCTGCTACTACACAGTAGTAAGGTATTTATCTCTGATACATTTTAC 4576
Qy 484 TCTCCGAGCGCGAAATGTTTCGAGTGTATGAAAAATACGCAATTCCTCGCGTGGAAATG 543
Db 4577 AATGCCGATCCAACTTCAATGATGCTTTGGAAAAAATGGGTATTTTAGGTATCGAAATG 4636
Qy 544 GAAGCGCTGTATCTACGGGCTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATCTGC 603
Db 4637 GAATCAGCTGTTTATATTTAAATGCGATTTCATGCTGTGTAATAAAGCACTTGGTATTTTC 4696
Qy 604 ACCGTATCTGACCATCCGACCTCACGAGCAGACCACTCGCGCTGAGCGTGCAGACTACC 663
Db 4697 ACAGTAAGTATCATATTTTACGTGACGAGCACTACTACACCTGAGGAAGCTCAAAATTTCA 4756
Qy 664 TTCAACGACATGATCAAAATCGCACTGGAATCCG 697

Db 4757 TTATACAAATGATGGAAATCGCTTTTAGAAATCG 4790
RESULT 11
US-09-583-110-37
Sequence 37, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
FILE REFERENCE: PATH00-07A
CURRENT APPLICATION NUMBER: US/09/583,110
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 37
LENGTH: 711
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-583-110-37

Query Match 28.4%; Score 204.2; DB 4; Length 711;
Best Local Similarity 57.0%; Pred. No. 6.9e-57;
Matches 393; Conservative 0; Mismatches 293; Indels 3; Gaps 1;
Qy 13 CACATTAATGCAGAAATGGCGATTTCCGCTGACGTAGTTTTTGATGCCAGGCGACCCGCTG 72
Db 10 CATATTCGCTGCTACGAGGGTGAATTTGCTGATAAAATTTCTTCTCTGGGGATCCTCTT 69
Qy 73 CGTCGAGATATATTCGTGAACCTTTCTTTGAAGATCCCGTGAAGTGAACAAAGTTTCGC 132
Db 70 CGTCTAAGTTTATTCGGGAGAAATTTCTTGGTGATGCTGTTTGTAAACGAAATGCGT 129
Qy 133 GGTATGCTGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTAATGGGTCAACCGT 192
Db 130 AACATGTTGGTACATCTGTACTTACAGGGTCAACCGTGTATCTGTCATGCGGAAGTGG 189
Qy 193 GTTGGTATCCGCTCCTCTCATCTACACCAAGAACTGATCACCGATTTTCGGGCTGAAG 252
Db 190 ATGGGAATGCCATCTATTTTCGATTTATGCGCGTGAGTTTAACTCGTAGACTACGGTGTGA 249
Qy 253 AAAATTAATCCGCTGGGTTCTGTGGCGCAGTTCTGCGGACGTAAACCTGCGGACGCTC 312
Db 250 AAATTTGATTCGTGTGGAACTGCGAGGTTCTTTGAATGAAGAGGTTTCATGTTTCGTGA 309
Qy 313 GTTATCGGTATGGTGGCTGCGACCGATTTCCAAAGTTTAAACCGCATCCGTTTAAAGACCAT 372
Db 310 GTTTTGGCGGAGCGGCTGCAACCAACTCAACATCGTTCTGTAATGACTGCGCCACAGTAC 369
Qy 373 GACTTTCCCGCTATCGCTGACTTCCACATGCTGGGTAAACGAGTAGATGCAAGCTAAAGCA 432
Db 370 GATTTTCCACAAATTTGCTAGCTTTTGTGTTGATTAAGGCTTACCATATCGCCAAAGAA 429
Qy 433 CTGGGTATGATGCTCGCGTGGGTAACTGTTTCTCCGCTGACCTGTTCTACTCTCCGAC 492
Db 430 CTTGGTATGACTACTCACTGCTGGGAAAGCTTTTGTTCATCTGATGTCTTTTACTCAAAAT 489
Qy 493 GCGCAATGTTTCGACGCTGATGAAAAATACGGCAATTTCTCGCGCTGGAATCGAAAGCGCT 552
Db 490 TTTGAAAGAAATATCGAGCTTGGTAAATGGGAATCAAGGCTGGAATGGAAGACGACGA 549
Qy 553 GGTATCTACGCGCTGCTGAGAAATTTTGGCGGAAAGCCCTGACCATCTGCAACCGCTATCT 612
Db 550 GCTCTTTTACTATCTTGTGCTGCCAATACCATGTTGATGCGCTAGCTATCATCACTCTCT 609
Qy 613 GACCACATCCGACTC---ACGACGACCACTCCGCTGAGCGCTCAGACTACCTTCAAC 669

Db 610 GATAGCTTGCTCAATCCAGACGAGAACACAACTGCAGAGAACGTCAAAATACCTTCACT 669
QY 670 GACATGATCAAAATCGACTGGATCCGT 698
Db 670 GATATGATGAAGGTGGTTTGGAAACCTT 698

RESULT 12

US-09-107-433-1967
; Sequence 1967, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1967:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 714 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...714
; SEQUENCE DESCRIPTION: SEQ ID NO: 1967:
US-09-107-433-1967

Query Match 28.48; Score 204.2; DB 4; Length 714;
Best Local Similarity 57.08; Pred. No. 6:9e-57;
Matches 393; Conservative 0; Mismatches 293; Indels 3; Gaps 1;

QY 13 CACATTAATGAGAAATGGCGATTTTCGCTGACGTAGTTTTGATGCCAGCGCACCGCTG 72
Db 13 CATATTGCTGCTCAGCAGGGTGAATTCGTGATAAATCTTCTCTCGGGATCCTCTT 72
QY 73 CGTGGCAAGTATATGCTGAAATTTCTTGAAGATCCCGTGAAGTGAACACGTTCCG 132
Db 73 CGTGGCAAGTATATGCTGAAATTTCTTGAAGATCCCGTGAAGTGAACACGTTCCG 132

QY 133 GGATATGCTGGCTTACCGGTACTTAAAGGCCGCAAAATTTCCGTAAATGGTGCAGGT 192
Db 133 AACATGTTTGGTTACACTGGTACTTACAGGGTCACGGTGTAICTGTCTATGGGAACGG 192
QY 193 GTTGGTATCCGTCCTGCTCCATCTACACAAAGAACTGATACCCGATTTCCGGCGGAAG 252
Db 193 ATGGGAATGCCATCTATTTCGATTTATGCGCGTGAGTTAATCGTAGACTACGGTGTGAAG 252
QY 253 AAAATTTATCGCGTGGTTCCTGTGGCGCAGTTCTGCCGCACGTAAACATGCCGCGAGCT 312
Db 253 AAATTCATTCGTGGGAACCTGCAGGTTCTTTGAATGAAGAGGTTTCATGTTCTGTGAATTA 312
QY 313 GTTATCGTATGGTGCCTGCACCGGATTCAAAAGTTTAAACCGCATCCGTTTAAAGACCAT 372
Db 313 GTTTTGGCGCAGCGGCTGCAACCACTCAACATCGTTGTAATGACTGGCCACAGTAC 372
QY 373 GACTTTGGCGCTATCGCTGACTTTTCGACATGGTTCGTAACGCGAGTAGATGCGACTAAAGCA 432
Db 373 GATTTTCCCAAAATTCGTAGCTTTGATTTGCTTGAATAAAGCCTTACCATATCGCCAAAGAA 432
QY 433 CTGGGTATTGATGCTCGCTGGGTAACTGTTCTCCGCTGACCTGTTCTACTCTCCGGAC 492
Db 433 CTGGGTATGACTACTCACGTTGGGAACGTTTGTCTCATCTGATGTCCTTTACTCAAATTAC 492
QY 493 GCGAAATGTTGACCGTGTATGAAAAAATACGGCATTTCTCGGCGTGGAAATGGAAGCGGT 552
Db 493 TTTGAAAAGAAATATCAGCTTGGTAAATGGGAGTCAAGGCTGTGGAATGGAAGCAGCA 552
QY 553 GGTATCTAGCGGTGCTGCGAGAAATTTGGCGGAAAGCCCTGACCATCTGCAACCGTATCT 612
Db 553 GCTCTTTACTATCTTCTGCTGCCAATACCATGTTGATGCGCTAGCTATCATGACCATCTCT 612
QY 613 GACCACATCGGACTC---ACGAGCAGACCACTGCGCTGAGCGTGCAGTCACTCTTCAAC 669
Db 613 GATAGCTTGGTCAATCCAGCAGAAACACAACTGCAGAAAGACGTCAAAATACCTTCACT 672
QY 670 GACATGATCAAAATCGCACTGGAATCCGT 698
Db 673 GATATGATGAAGTTGGTTTGGAAACCTT 701

RESULT 13

US-08-961-527-128
; Sequence 128, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:

```
;
;
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 128:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-128

Query Match      28.4%; Score 204.2; DB 3; Length 13440;
Best Local Similarity 57.0%; Pred. No. 4e-56;
Matches 393; Conservative 0; Mismatches 293; Indels 3; Gaps 1;

Qy 13 CACATTAATGCAGAAATGGCGATTTTCGCTGACGTAGTTTGTATGCCAGGCGACCCGCTG 72
Db 10453 CATATTGCTGCTCAGCAGGGTGAATTCGTGATAAAATTCCTTCCTGGGATCCTCTT 10512

Qy 73 CGTCGGAAGTATATGCTGAAACTTTCTTGAAGATGCCGTGAAGTGAACAACTGTCGC 132
Db 10513 CGTCTAAGTTTATTCGGGAAATTCCTTGTATGATGCTGTTGTTTAAACGAAGTGGT 10572

Qy 133 GGTATGCTGGCTTCACCGGTACTTACAAAGCGCAAAATTTCCGTAATGGTCAACGGT 192
Db 10573 AACATGTTGTTACATGCTGACTTACAAAGGTCACTGTGATCTGTCATGGAACTGGG 10632

Qy 193 GTTGGTATCCGCTCCTGCTCATCTACACCAAGAACTGATCAACGATTTTCGGCGTGAAG 252
Db 10633 ATGGGAATGCCATCTATTTCGATTTATCGCGGTGAGTTAATCGTAGACTACGGTGTGAG 10692

Qy 253 AAAATTATCCGCGTGGTTCCTGTGGCGCAGTTCCTGGCGCAGCTGAAACCTGCGGAGTC 312
Db 10693 AAATTGATTCGTGTGGAACTGCAGGTTCTTTGAATGAAGAGGTTTCATGTTCTGTGAATTA 10752

Qy 313 GTTATCGGTATGGTGTCTGCACCGATTCCAAAGTTAACCGCATCCGTTTAAAGACCAT 372
Db 10753 GTTTGGCGGAGCGGCTGCAACCACTCAAAACATCGTTTCGTAATGACTGGCCACAGTAC 10812

Qy 373 GACTTTGCCGCTATCGCTGACTTCGACATGTTGGTGAACGAGTAGATGCAGCTTAAAGCA 432
Db 10813 GATTTTCCACAATTTGCTAGCTTTGATTTGTTGTTGATTAAGCTACCATATCGCCAAAAA 10872

Qy 433 CTGGGTATGATGCTCGCGTGGGTAACTGTTCTTCGCGTGACCTGTTCTACTCTCCGGAC 492
Db 10873 CTTGGTATGACTACTCAGTTGGGAAGCTTTTGTCTATGCTGATGCTTTTACTCAAATTAC 10932

Qy 493 GCGCAATGTTGACGCTGATGGAATAATACGGCATTCCTCGCGTGGAAATGGAAGCGGCT 552
Db 10933 TTTGAAAGAAATATCGAGCTTGGTAAATGGGAGTCAAGGCTGTGGAATGGAAGCAGCA 10992

Qy 553 GGTATCTACGGCGTCTGTCAGAAATTTGGCGGAAAGCCCTGACCATCTGCACCGTATCT 612
Db 10993 GCTCTTACTATCTTCTGTCGCCAATACCATGTTGATGGCTAGCTATCATGACCATCTCT 11052

Qy 613 GACCACATCCGCATC---ACGACGACACACTCCGCGTGAGCGTGACACTACCTTCAAC 669
Db 11053 GATAGCTTGGTCAATCCAGACGAGACACAACTGCAGAAAGAACTCAAAATACCTTCACT 11112

Qy 670 GACATGATCAAAATCGCACTGGAATCCGT 698
Db 11113 GATATGATGAGGTTGGTTTGGAAACCTT 11141

RESULT 14
US-09-107-532A-1185
; Sequence 1185, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
```

```
;
;
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...735
; SEQUENCE DESCRIPTION: SEQ ID NO: 1185:
; US-09-107-532A-1185

Query Match      27.6%; Score 198.6; DB 4; Length 735;
Best Local Similarity 55.6%; Pred. No. 5e-55;
Matches 381; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

Qy 13 CACATTAATGCAGAAATGGCGATTTTCGCTGACGTAGTTTGTATGCCAGGCGACCCGCTG 72
Db 37 CATATTGAAGCGAAAGAGCGGAGATCGCCGATTAAGATTTTACTTCCGGGAGATCCTTTTG 96

Qy 73 CGTCGGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGTGAAGTGAACAAAGTTTCGC 132
Db 97 CGAGCTAAATATATTGGCGAGACTTTTGGAGATCCAGTTTGTACATCAGGTAGCG 156

Qy 133 GGTATGCTGGCTTCACCGTACTTACAAAGCGCAAAATTTCCGTAATGGGTACCGGT 192
Db 157 GGTATGTTAGCTTTTACCGCAAAATATAAAGGAGAACGTATTTCTGTTCAAGGAACAGGA 216

Qy 193 GTTGGTATCCGCTCCTGCTCCATCTACCAAGAACTGATCACCAGTTTCGGCGTGAAG 252
Db 217 ATGGGAATGCCATCAGCCACCATCTATGCCCATGAATTTGATCCAGTCGATGGTGAAG 276

Qy 253 AAAATTATCCGCGTGGTTCCTGTGGCGCAGTTCTCGCGCACGTAAACCTCGCGACGTC 312
Db 277 AAATTGATTGAGTAGGTACATCGGTGTCATTGTCAAAGACGTCCTATGTCGTGATTTA 336

Qy 313 GTTATCGGTATGGTGCCTGCACCGATTCCAAAGTTAAACCGCATCCGTTTTAAAGACCAT 372
Db 337 GTACTAGCAAAAGGTGCAGCGACCGAGTTCTTCGATGATCGAAAAAATTTTCAAAGCCTTT 396

Qy 373 GACTTTCGCGTATCGCTGACTTCGACATGGTGGTGAACGAGTAGATGCAGCTAAAGCA 432
```

Db 397 CATTTTCTCCCAATCAGTGATTTCAATCTATTACTGAAAGCATATGAGATAGCTAAAGAA 456
QY 433 CTGGTATTGATGCTCGCGGGGTAACTCTTCTCGCTGACCTGTCTACTCTCCGGAC 492
Db 457 AAGGGCTATACGGTTCAATGATAGGAATGTTCTTTCGGAAGATTCTTCTTACAAGATGAT 516
QY 493 GGGGAAATGTTGACACGTGATGGAAAAATACCGCATTTCTCGCGCTGGAAAAATGGAAGCGGCT 552
Db 517 TTGACAGAGACGTTCCAAATTAGCTGAACTTGGCGTTTAGAGTGGAAATGGAGCAGCG 576
QY 553 GGTATCTACCGCGTGGCTGACGAATTTGGCGGAAAGCCCTGACCATCTGCAACCGTATCT 612
Db 577 GCTCTTTATTACTTAGCGCGGAAATACCATGTGCAGACATTAAAGTTTGTGATGACCGTTAGC 636
QY 613 GACCATCCGCACTCACAGCAGACACCTGCGCTGAGCGTCAAGTACTACCTCAACGAC 672
Db 637 GACCATTTGATACCCGGTGAAGAAACAAACAGCTGCAGAACGTCAATCTACGTTCAATGAA 696
QY 673 ATGATCAAAATCGCACTGGAATCCG 697
Db 697 ATGATCGAGTAGGATTGGAAACAG 721

RESULT 15

US-09-710-279-3511
; Sequence 3511, Application US/09710279
; Patent No. 6703492

; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US

; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09

; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09

; NUMBER OF SEQ ID NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 3511

; LENGTH: 2962

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic

; OTHER INFORMATION: nucleic acid sequence

US-09-710-279-3511

Query Match 25.1%; Score 180.8; DB 4; Length 2962;
Best Local Similarity 53.9%; Pred. No. 8.6e-49;
Matches 371; Conservative 0; Mismatches 317; Indels 0; Gaps 0;

QY 6 TACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTGATGCCAGCGCA 65
Db 1640 TACACCTCACATTAACCCAAATGGAGTTAAATAGCTTAAACTGTATTAAATGCCAGCGCA 1699
QY 66 CCGCGTGGTGCAGAAATATATGCTGAAACTTTCTTGAAGATGCCCGTGAAGTGAACAA 125
Db 1700 CCCATTACGTGCTCAATATATGCTGAAATTTTGTAGAAATGTTGAACAATTCATATC 1759
QY 126 CGTTCGGGTATGCTGGGCTTCCCGGTACTTACAAAGCGCGCAAAATTTCCGTAATGGG 185
Db 1760 AGTACGTAATATGTTGGTTTACACAGGAACCTTATAAGGCAAAAGATTTCTGTGATGG 1819
QY 186 TCACGGTGTGGTATCCCGTCCCTGCTCCATCTACACCAAGAACTGATCACCGATTTCCG 245
Db 1820 CTCCTGGATGGGAATTCGAAGTATGGTATTATTTCTATGAACTTACCAATTCCTTGA 1879
QY 246 CGTGAAGAAATATCCGCGTGGGTTCTGTGGCGAGTTTCTCCGCGACGTAAAACTGGG 305
Db 1880 TGTGGATACAATCATTCGCGTAGGTTTCATGTGGCGCTCTTCAAGAAAGATGTCAATTTATA 1939
QY 306 CGACGTCGTTATCGGTATGGGTGCTGCCCGGATTCCAAAGTTAACCGGATCCGTTTTAA 365
Db 1940 TGATGTGATTATTGCAACAGCCGCTTCCCAAACTCAAATTTATGTTGATCAATTTAATAT 1999

QY 366 AGACCATGACTTTGCGCTATCGCTGACTTCGACATGGTGCCTAAACGACGTAGATGCAGC 425
Db 2000 ACCAGGTCACTTTGCGCTATTGCTGATATATCTAGTAGCTAAAGCTAAGAAAGCTGC 2059
QY 426 TAAAGCACCTGGGTATTGATGCTCGCGTGGGTAACTGTCTCCGCTGACCTGTCTTACTC 485
Db 2060 TGATGAAATTTGGTGGCCATATCATGTAGGTAAAGTTTTGTCTCATCTGATACATTTTATAA 2119
QY 486 TCCGGAGGGGAAATGTTTCGACGTGATGGAAAAATACGGCATTTCTCGGCGTGGAAATGGA 545
Db 2120 TGCTGATTCAACTTTTAAATGATTTCATGGAATAAGATGGGTATCTTAGGTATCGAAATGGA 2179
QY 546 AGCGGCTGGTATCTACGGGCTGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATCTGCAC 605
Db 2180 ATCTGCAAGTCTATATTTAAATGCCATCCATGCAAAATAAAAGGCTTTAGGTATTTTAC 2239
QY 606 CGTATCTGACCACTCCGCACTCACGAGCAGACCACTGCGCGTGAAGCGTCAGACTACTCTT 665
Db 2240 AGTAAAGTGTATCATATTCTTAAGAGATGAAGCCACTAGTGCAGAAAGAGAGACAAACATCAT 2299
QY 666 CAACGACATGATCAAAATCGCACTGGAA 693
Db 2300 TACTCAATGATGGAATAATAGCGCTCGAA 2327

Search completed: July 22, 2005, 22:31:51

Job time : 242 secs

Result	Query	Score	Length	DB	ID	Description
No.	Match					

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80																				

ALIGNMENTS

```
QY 61 GGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 120
Db |||
QY 61 GGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 120
Db |||
QY 121 AACAAAGCTTCCGCGTATGCTGGGCTTCAACCGGTATCTTCAAAAGCCGCAAAATTTCCGTA 180
Db |||
QY 121 AACAAAGCTTCCGCGTATGCTGGGCTTCAACCGGTATCTTCAAAAGCCGCAAAATTTCCGTA 180
Db |||
QY 181 ATGGGTACACCGTGTGGTATCCCGTCTCGCTCCATCTACACCAAGAACTGATCACCGAT 240
Db |||
QY 181 ATGGGTACACCGTGTGGTATCCCGTCTCGCTCCATCTACACCAAGAACTGATCACCGAT 240
Db |||
QY 241 TTGGCGCTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAAAA 300
Db |||
QY 241 TTGGCGCTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAAAA 300
Db |||
QY 301 CTGGCGACGTCGTTATCGGTATGGGTGCTGACCGATTTCAAAGTTAAACCGCATCCGT 360
Db |||
QY 301 CTGGCGACGTCGTTATCGGTATGGGTGCTGACCGATTTCAAAGTTAAACCGCATCCGT 360
Db |||
QY 361 TTTAAAGACCATGACTTTGGCGTATCGGTATGGGTGCTGACCGATTTCAAAGTTAAACCGTAGAT 420
Db |||
QY 361 TTTAAAGACCATGACTTTGGCGTATCGGTATGGGTGCTGACCGATTTCAAAGTTAAACCGTAGAT 420
Db |||
QY 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
Db |||
QY 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
Db |||
QY 481 TACTCTCCGACGCGGAAATTTGCGAGTGTGAGTGAAGAAATTAACCGCATTTCTCGGCGTGA 540
Db |||
QY 481 TACTCTCCGACGCGGAAATTTGCGAGTGTGAGTGAAGAAATTAACCGCATTTCTCGGCGTGA 540
Db |||
QY 541 ATGGAGCGGCTGCTATCTACCGCGTGGTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db |||
QY 541 ATGGAGCGGCTGCTATCTACCGCGTGGTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db |||
QY 601 TGACCCGTATCTGACCATCTCCGACCTCAGCAGCAGACCACTGCGCGTGAAGCGTCAAGCT 660
Db |||
QY 601 TGACCCGTATCTGACCATCTCCGACCTCAGCAGCAGACCACTGCGCGTGAAGCGTCAAGCT 660
Db |||
QY 661 ACCTTCAACGACATGATCAAAATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCG 720
Db |||
QY 661 ACCTTCAACGACATGATCAAAATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCG 720
Db |||
```

RESULT 2

```
US-10-891-096-3
; Sequence 3, Application US/10891096
; Publication No. US20050074857A1
; GENERAL INFORMATION:
; APPLICANT: Araki, Tadaishi
; APPLICANT: Miyake, Hitoki
; APPLICANT: Oikawa, Toshihiro
; TITLE OF INVENTION: Method for Producing a Pyrimidine Nucleoside Compound
; TITLE OF INVENTION: and a New Pyrimidine Nucleoside Compound
; FILE REFERENCE: 018765-161
; CURRENT APPLICATION NUMBER: US/10/891,096
; CURRENT FILING DATE: 2004-07-15
; PRIOR APPLICATION NUMBER: JP 2003-199175
; PRIOR FILING DATE: 2003-07-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-891-096-3
```

Query Match 99.6%; Score 716.8; DB 21; Length 720;
Best Local Similarity 99.7%; Pred. No. 3e-228;
Matches 716; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGCTAGTTTGTGATGCCA 60

```
Db |||
QY 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTTGCTGACGCTAGTTTGTGATGCCA 60
Db |||
QY 61 GGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 120
Db |||
QY 61 GGGACCCGCTGCGTGCAGAGTATATGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 120
Db |||
QY 121 AACAAAGCTTCCGCGTATGCTGGGCTTCAACCGGTATCTTCAAAAGCCGCAAAATTTCCGTA 180
Db |||
QY 121 AACAAAGCTTCCGCGTATGCTGGGCTTCAACCGGTATCTTCAAAAGCCGCAAAATTTCCGTA 180
Db |||
QY 181 ATGGGTACACCGTGTGGTATCCCGTCTCGCTCCATCTACACCAAGAACTGATCACCGAT 240
Db |||
QY 181 ATGGGTACACCGTGTGGTATCCCGTCTCGCTCCATCTACACCAAGAACTGATCACCGAT 240
Db |||
QY 241 TTGGCGCTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAAAA 300
Db |||
QY 241 TTGGCGCTGAAGAAATTTATCCGCGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAAAA 300
Db |||
QY 301 CTGGCGACGTCGTTATCGGTATGGGTGCTGACCGATTTCAAAGTTAAACCGCATCCGT 360
Db |||
QY 301 CTGGCGACGTCGTTATCGGTATGGGTGCTGACCGATTTCAAAGTTAAACCGCATCCGT 360
Db |||
QY 361 TTTAAAGACCATGACTTTGGCGTATCGGTATGGGTGCTGACCATGTTCAAGTTAAACCGTAGAT 420
Db |||
QY 361 TTTAAAGACCATGACTTTGGCGTATCGGTATGGGTGCTGACCATGTTCAAGTTAAACCGTAGAT 420
Db |||
QY 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
Db |||
QY 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTTC 480
Db |||
QY 481 TACTCTCCGACGCGGAAATTTGCGAGTGTGAGTGAAGAAATTAACCGCATTTCTCGGCGTGA 540
Db |||
QY 481 TACTCTCCGACGCGGAAATTTGCGAGTGTGAGTGAAGAAATTAACCGCATTTCTCGGCGTGA 540
Db |||
QY 541 ATGGAGCGGCTGCTATCTACCGCGTGGTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db |||
QY 541 ATGGAGCGGCTGCTATCTACCGCGTGGTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db |||
QY 601 TGACCCGTATCTGACCATCTCCGACCTCAGCAGCAGACCACTGCGCGTGAAGCGTCAAGCT 660
Db |||
QY 601 TGACCCGTATCTGACCATCTCCGACCTCAGCAGCAGACCACTGCGCGTGAAGCGTCAAGCT 660
Db |||
QY 661 ACCTTCAACGACATGATCAAAATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCG 720
Db |||
QY 661 ACCTTCAACGACATGATCAAAATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCG 720
Db |||
```

RESULT 3

```
US-09-891-865A-12
; Sequence 12, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 3031
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: udp and deod
; OTHER INFORMATION: cloned into pGM746 without upstream ptac promoter
US-09-891-865A-12
```

Query Match 99.6%; Score 716.8; DB 10; Length 3031;

; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patent in ver. 2.1
; SEQ ID NO 13
; LENGTH: 3128
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: deoD cloned
; OTHER INFORMATION: downstream ptac promoter
US-09-891-865A-13

Query Match 99.6%; Score 716.8; DB 10; Length 3128;
Best Local Similarity 99.7%; Pred. No. 5.8e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGCTGACGTAGTTTGTATGCCA 60
DB 198 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGCTGACGTAGTTTGTATGCCA 257

QY 61 GGGACCCGCTGCGTGGGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGCGTGAAGTG 120
DB 258 GGGACCCGCTGCGTGGGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGCGTGAAGTG 317

QY 121 AACAAAGTTTCGCGGTATGCTGGGCTTCACCGGTACTTTACAAAGCCGCAAAATTTCCGTA 180
DB 318 AACAAAGTTTCGCGGTATGCTGGGCTTCACCGGTACTTTACAAAGCCGCAAAATTTCCGTA 377

QY 181 ATGGGTCAACGCTGTTGGTATCCCGTCTCGCTCCATCTACACAAAGAACTGTACCCGAT 240
DB 378 ATGGGTCAACGCTGTTGGTATCCCGTCTCGCTCCATCTACACAAAGAACTGTACCCGAT 437

QY 241 TTGGCGTGAAGAAATTTATCCGGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAA 300
DB 438 TTGGCGTGAAGAAATTTATCCGGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAA 497

QY 301 CTGCGGACGCTGTTATCCGCTATGCTGGGCTGCTGACCGATTCCAAAGTTAACCCGATCCGT 360
DB 498 CTGCGGACGCTGTTATCCGCTATGCTGGGCTGCTGACCGATTCCAAAGTTAACCCGATCCGT 557

QY 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGGTAAACGAGTAGAT 420
DB 558 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGGTAAACGAGTAGAT 617

QY 421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
DB 618 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 677

QY 481 TACTCTCCGACCGCGAAATTTTCGACGTGATGGAATAATACCGCATTTCTCGCGGTGGAA 540
DB 678 TACTCTCCGACCGCGAAATTTTCGACGTGATGGAATAATACCGCATTTCTCGCGGTGGAA 737

QY 541 ATGGAAGCGCTGGTATCTACCGCGTGGGTAAACCTGTTCTCCGCTGACCTGACCATC 600
DB 738 ATGGAAGCGCTGGTATCTACCGCGTGGGTAAACCTGTTCTCCGCTGACCTGACCATC 797

QY 601 TGACCGGTATCTGACACCATCCGCACTACAGAGCAGACCACTGCGCGTGAAGCTGACACT 660
DB 798 TGACCGGTATCTGACACCATCCGCACTACAGAGCAGACCACTGCGCGTGAAGCTGACACT 857

QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
DB 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 917

RESULT 6

US-11-036-497-13
; Sequence 13 Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona

; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucleic acid analogues thereof
; FILE REFERENCE: 02901/000J410-USO
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 13
; LENGTH: 3128
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: deoD cloned downstream ptac promoter
US-11-036-497-13

Query Match 99.6%; Score 716.8; DB 24; Length 3128;
Best Local Similarity 99.7%; Pred. No. 5.8e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGCTGACGTAGTTTGTATGCCA 60
DB 198 ATGGCTACCCACACATTAATGCAGAAATGGGCGATTTCGCTGACGTAGTTTGTATGCCA 257

QY 61 GGGACCCGCTGCGTGGGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGCGTGAAGTG 120
DB 258 GGGACCCGCTGCGTGGGAAGTATATTGCTGAAACTTTCTTGAAGATGCCGCGTGAAGTG 317

QY 121 AACAAAGTTTCGCGGTATGCTGGGCTTCACCGGTACTTTACAAAGCCGCAAAATTTCCGTA 180
DB 318 AACAAAGTTTCGCGGTATGCTGGGCTTCACCGGTACTTTACAAAGCCGCAAAATTTCCGTA 377

QY 181 ATGGGTCAACGCTGTTGGTATCCCGTCTCGCTCCATCTACACAAAGAACTGTACCCGAT 240
DB 378 ATGGGTCAACGCTGTTGGTATCCCGTCTCGCTCCATCTACACAAAGAACTGTACCCGAT 437

QY 241 TTGGCGTGAAGAAATTTATCCGGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAA 300
DB 438 TTGGCGTGAAGAAATTTATCCGGTGGGTTCTGTGGCGCAGTTCTGCGCGCAGTAA 497

QY 301 CTGCGGACGCTGTTATCCGCTATGCTGGGCTGCTGACCGATTCCAAAGTTAACCCGATCCGT 360
DB 498 CTGCGGACGCTGTTATCCGCTATGCTGGGCTGCTGACCGATTCCAAAGTTAACCCGATCCGT 557

QY 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGGTAAACGAGTAGAT 420
DB 558 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGGTAAACGAGTAGAT 617

QY 421 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
DB 618 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 677

QY 481 TACTCTCCGACCGCGAAATTTTCGACGTGATGGAATAATACCGCATTTCTCGCGGTGGAA 540
DB 678 TACTCTCCGACCGCGAAATTTTCGACGTGATGGAATAATACCGCATTTCTCGCGGTGGAA 737

QY 541 ATGGAAGCGCTGGTATCTACCGCGTGGGTAAACCTGTTCTCCGCTGACCTGACCATC 600
DB 738 ATGGAAGCGCTGGTATCTACCGCGTGGGTAAACCTGTTCTCCGCTGACCTGACCATC 797

QY 601 TGACCGGTATCTGACACCATCCGCACTACAGAGCAGACCACTGCGCGTGAAGCTGACACT 660
DB 798 TGACCGGTATCTGACACCATCCGCACTACAGAGCAGACCACTGCGCGTGAAGCTGACACT 857

QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
DB 917 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 917

Db 858 ACCTTCAACGACATGATCAAAATCCCACTGGAAATCCGTTCTGCTGGGCGATAAAGAGTAA 917

RESULT 7
US-09-891-865A-3
; Sequence 3, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: 99DC24E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3383
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; NAME/KEY: gene
; LOCATION: (231)..(960)
; OTHER INFORMATION: deod
US-09-891-865A-3

Query Match 99.6%; Score 716.8; DB 10; Length 3383;
Best Local Similarity 99.7%; Pred. No. 6e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 60
Db 240 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 299

Qy 61 GGCACCCGCTGCTGCGGATGCTGGCTTACCGGTACTTACAAAGGCGCGAAATTTCCGTA 120
Db 300 GGCACCCGCTGCTGCGGATGCTGGCTTACCGGTACTTACAAAGGCGCGAAATTTCCGTA 359

Qy 121 AACAAAGTTCGCGGTATGCTGGCTTACCGGTACTTACAAAGGCGCGAAATTTCCGTA 180
Db 360 AACAAAGTTCGCGGTATGCTGGCTTACCGGTACTTACAAAGGCGCGAAATTTCCGTA 419

Qy 181 ATGGGTACCGGTGTTGGTATCCCGTCCCTGCTTACCAAGGCGCGAAATTTCCGTA 240
Db 420 ATGGGTACCGGTGTTGGTATCCCGTCCCTGCTTACCAAGGCGCGAAATTTCCGTA 479

Qy 241 TTCGCGGTGAAGAAATTTATCCGCGTGGGTTCCTGGAAGATGCCCGTGAAGTG 300
Db 480 TTCGCGGTGAAGAAATTTATCCGCGTGGGTTCCTGGAAGATGCCCGTGAAGTG 539

Qy 301 CTGCGCGACGTGCTGTTATCGGTATCGGTGCTTACCAAGGCGCGAAATTTCCGTA 360
Db 540 CTGCGCGACGTGCTGTTATCGGTATCGGTGCTTACCAAGGCGCGAAATTTCCGTA 599

Qy 361 TTTAAAGACCATGATCTTTGCGGCTATCGCTGACATGGTGGCGTAAACGCGATAGAT 420
Db 600 TTTAAAGACCATGATCTTTGCGGCTATCGCTGACATGGTGGCGTAAACGCGATAGAT 659

Qy 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTC 480
Db 660 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTC 719

Qy 481 TACTCTCCGACGCGGAAATGTTTGAACCTGATGGAATAATCGGATTTCTCGGCGTGAA 540
Db 720 TACTCTCCGACGCGGAAATGTTTGAACCTGATGGAATAATCGGATTTCTCGGCGTGAA 779

Qy 541 ATGGAAGCGGTGCTGATCTACCGGTGCTGAGAAATTTGGGCGGAAGCCCTGACCATC 600
Db 780 ATGGAAGCGGTGCTGATCTACCGGTGCTGAGAAATTTGGGCGGAAGCCCTGACCATC 839

Qy 601 TGCACCGTATCTGACCACTCCGCACTCAACGAGAGACCACTGCGGCTGAGCGTCAGACT 660
Db 840 TGCACCGTATCTGACCACTCCGCACTCAACGAGAGACCACTGCGGCTGAGCGTCAGACT 899

Qy 661 ACCTTCAACGACATGATCAAAATCCCACTGGAAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db 900 ACCTTCAACGACATGATCAAAATCCCACTGGAAATCCGTTCTGCTGGGCGATAAAGAGTAA 959

RESULT 8
US-11-036-497-3
; Sequence 3, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; FILE REFERENCE: 02901/000J410-US0
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 3383
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid
; NAME/KEY: gene
; LOCATION: (231)..(960)
; OTHER INFORMATION: deod
US-11-036-497-3

Query Match 99.6%; Score 716.8; DB 24; Length 3383;
Best Local Similarity 99.7%; Pred. No. 6e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 60
Db 240 ATGGCTACCCACACATTAATGCAGAAATGGCGGATTCGCTGACGTAGTTTGTATGCCA 299

Qy 61 GGCACCCGCTGCTGCGGATGCTGGCTTACCGGTACTTACAAAGGCGCGAAATTTCCGTA 120
Db 300 GGCACCCGCTGCTGCGGATGCTGGCTTACCGGTACTTACAAAGGCGCGAAATTTCCGTA 359

Qy 121 AACAAAGTTCGCGGTATGCTGGCTTACCGGTACTTACAAAGGCGCGAAATTTCCGTA 180
Db 360 AACAAAGTTCGCGGTATGCTGGCTTACCGGTACTTACAAAGGCGCGAAATTTCCGTA 419

Qy 181 ATGGGTACCGGTGTTGGTATCCCGTCCCTGCTTACCAAGGCGCGAAATTTCCGTA 240
Db 420 ATGGGTACCGGTGTTGGTATCCCGTCCCTGCTTACCAAGGCGCGAAATTTCCGTA 479

Qy 241 TTCGCGGTGAAGAAATTTATCCGCGTGGGTTCCTGGAAGATGCCCGTGAAGTG 300
Db 480 TTCGCGGTGAAGAAATTTATCCGCGTGGGTTCCTGGAAGATGCCCGTGAAGTG 539

Qy 301 CTGCGCGACGTGCTGTTATCGGTATCGGTGCTTACCAAGGCGCGAAATTTCCGTA 360
Db 540 CTGCGCGACGTGCTGTTATCGGTATCGGTGCTTACCAAGGCGCGAAATTTCCGTA 599

Qy 361 TTTAAAGACCATGATCTTTGCGGCTATCGCTGACATGGTGGCGTAAACGCGATAGAT 420

Db 600 TTTAAAGACCATGCTTTGCGGTATCGCTGACTTCGACATGGTGGGTAAACCGTAGAT 659
QY 421 GCAGCTAAAGCACTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCGCTGACCTGTTTC 480
Db 660 GCAGCTAAAGCACTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCGCTGACCTGTTTC 719
QY 481 TACTCTCCGACCGCGGAAATGTTTCGAGTGTGATGGAATAATACGGCATTTCTCGCGGTGAA 540
Db 720 TACTCTCCGACCGCGGAAATGTTTCGAGTGTGATGGAATAATACGGCATTTCTCGCGGTGAA 779
QY 541 ATGGAAGCGGTGATCTACCGCGTTCGCTGACAGATTTGCGCGGAAAGCCCTGACCATC 600
Db 780 ATGGAAGCGGTGATCTACCGCGTTCGCTGACAGATTTGCGCGGAAAGCCCTGACCATC 839
QY 601 TGCACCGTATCTGACCAATCCGCACTCAAGAGCAGACCACTCGCGCTGAGCGTCAAGT 660
Db 840 TGCACCGTATCTGACCAATCCGCACTCAAGAGCAGACCACTCGCGCTGAGCGTCAAGT 899
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 959
RESULT 9
US-09-891-865A-14
; Sequence 14, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: udp and deoD
; OTHER INFORMATION: cloned downstream ptac promoter
US-09-891-865A-14

Query Match 99.6%; Score 716.8; DB 10; Length 3934;
Best Local Similarity 99.7%; Pred. No. 6.5e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACATTAATGAGAAATGGCGATTTGCTGACGTAGTTTGTATGCCA 60
Db 198 ATGGCTACCCACACATTAATGAGAAATGGCGATTTGCTGACGTAGTTTGTATGCCA 257
QY 61 GCGACCGCTGCTGGAGATATGCTGAACCTTTCTTGAAGTCCCGTGAAGTG 120
Db 258 GCGACCGCTGCTGGAGATATGCTGAACCTTTCTTGAAGTCCCGTGAAGTG 317
QY 121 AACAAAGTTCCGGTATGCTGGCGTTCCACCGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db 318 AACAAAGTTCCGGTATGCTGGCGTTCCACCGTACTTACAAAGCGCGCAAAATTTCCGTA 377
QY 181 ATGGGTACCGGTGTTGGTATCCCGTCTGCTCCATCTTACAAAGAACCTGATCACCGAT 240
Db 378 ATGGGTACCGGTATGGTATCCCGTCTGCTCCATCTTACAAAGAACCTGATCACCGAT 437
QY 241 TTCCGGGTGAAGAAATTAATCCCGGTGGGTTCCTGTCGGCGACGTTCTGCGCGCACTAAA 300
Db 438 TTCCGGGTGAAGAAATTAATCCCGGTGGGTTCCTGTCGGCGACGTTCTGCGCGCACTAAA 497
QY 301 CTGCGGACGCTGTTATCGGTATGGGTGCTGCAACCGATTTCCAAAGTTAAACCGCATCCGT 360

Db 498 CTGCGGACGCTGCTTATCGGTATGGGTGCTGACCGATTTCAAAGTTAACCGCATCCGT 557
QY 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTTCGACATGGTGGGTAAACCGAGTAGAT 420
Db 558 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTTCGACATGGTGGGTAAACCGAGTAGAT 617
QY 421 GCAGCTAAAGCACTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
Db 618 GCAGCTAAAGCACTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 677
QY 481 TACTCTCCGACCGCGGAAATGTTTCGACGTGATCGAAAAATACGGCATTTCTCGCGGTGAA 540
Db 678 TACTCTCCGACCGCGGAAATGTTTCGACGTGATCGAAAAATACGGCATTTCTCGCGGTGAA 737
QY 541 ATGGAAGCGGTGATCTACCGCGTTCGCTGACAGATTTTGGCGCGGAAAGCCCTGACCATC 600
Db 738 ATGGAAGCGGTGATCTACCGCGTTCGCTGACAGATTTTGGCGCGGAAAGCCCTGACCATC 797
QY 601 TGCACCGTATCTGACCAATCCGCACTCAAGAGCAGACCACTCGCGCTGAGCGTCAAGT 660
Db 798 TGCACCGTATCTGACCAATCCGCACTCAAGAGCAGACCACTCGCGCTGAGCGTCAAGT 857
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 720
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAGTAA 917
RESULT 10
US-11-036-497-14
; Sequence 14, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; TITLE OF INVENTION: and modified analogues thereof
; FILE REFERENCE: 02901/000J410-USO
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 3934
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: udp and deoD cloned downstream ptac promoter
US-11-036-497-14

Query Match 99.6%; Score 716.8; DB 24; Length 3934;
Best Local Similarity 99.7%; Pred. No. 6.5e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCTACCCACACATTAATGAGAAATGGCGATTTGCTGACGTAGTTTGTATGCCA 60
Db 198 ATGGCTACCCACACATTAATGAGAAATGGCGATTTGCTGACGTAGTTTGTATGCCA 257
QY 61 GCGACCGCTGCTGGAGATATGCTGAACCTTTCTTGAAGTCCCGTGAAGTG 120
Db 258 GCGACCGCTGCTGGAGATATGCTGAACCTTTCTTGAAGTCCCGTGAAGTG 317
QY 121 AACAAAGTTCCGGTATGCTGGCGTTCCACCGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Db 318 AACAAAGTTCCGGTATGCTGGCGTTCCACCGTACTTACAAAGCGCGCAAAATTTCCGTA 377

Qy 181 ATGGGTACCGGTGTTGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 240
Db |||||
Qy 378 ATGGGTACCGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 437
Db |||||
Qy 241 TTCGGCGTGAAGAAATATCCCGTGGGTTCCTGTGGCGAGTTCGCCGCACTGAAA 300
Db |||||
Qy 438 TTCGGCGTGAAGAAATATCCCGTGGGTTCCTGTGGCGAGTTCGCCGCACTGAAA 497
Db |||||
Qy 301 CTGGCGCACTGCTTATCCGCTATGGGTGCTGCAACCGATTCCAAAGTTAACCGCATCCGT 360
Db |||||
Qy 498 CTGGCGCACTGCTTATCCGCTATGGGTGCTGCAACCGATTCCAAAGTTAACCGCATCCGT 557
Db |||||
Qy 361 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGCCTGAAAGCGCAGTAGAT 420
Db |||||
Qy 558 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGCCTGAAAGCGCAGTAGAT 617
Db |||||
Qy 421 GCAGCTAAAGCACATGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTTTC 480
Db |||||
Qy 618 GCAGCTAAAGCACATGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCGCGTGAACCTGTTTC 677
Db |||||
Qy 481 TACTCTCCGACGCGGAAATGTTCCGACTGATGAAATAACGGCATTCCTCGCGTGGAA 540
Db |||||
Qy 678 TACTCTCCGACGCGGAAATGTTCCGACTGATGAAATAACGGCATTCCTCGCGTGGAA 737
Db |||||
Qy 541 ATGAAAGCGGTGCTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db |||||
Qy 738 ATGAAAGCGGTGCTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 797
Db |||||
Qy 601 TGCACCGTATCTGACCATCCGCACTCAGAGCAGACCACTGCGCGTGAAGCTGAGCTCAGACT 660
Db |||||
Qy 798 TGCACCGTATCTGACCATCCGCACTCAGAGCAGACCACTGCGCGTGAAGCTGAGCTCAGACT 857
Db |||||
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db |||||
Qy 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 917
Db |||||

RESULT 11

US-09-891-865A-5
; Sequence 5, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891,865A
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5

; LENGTH: 4189
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; FEATURE:
; NAME/KEY: gene
; LOCATION: (243)..(1021)
; OTHER INFORMATION: udp
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1037)..(1766)
; OTHER INFORMATION: deod
US-09-891-865A-5

Query Match 99.6%; Score 716.8; DB 10; Length 4189;
Best Local Similarity 99.7%; Pred. No. 6.7e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGACAGAAATGGGCGATTTTCGTGACGTAGTTTGTGATGCCA 60
Db |||||
Qy 1046 ATGGCTACCCACACATTAATGACAGAAATGGGCGATTTTCGTGACGTAGTTTGTGATGCCA 1105
Db |||||
Qy 61 GGGACCCGCTGCGTGGCGAAGTATATTGCTGAAACTTTCCCTTGAAGATGCCCGTGAAGTG 120
Db |||||
Qy 1106 GGGACCCGCTGCGTGGCGAAGTATATTGCTGAAACTTTCCCTTGAAGATGCCCGTGAAGTG 1165
Db |||||
Qy 121 AACAACTGTCGGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
Db |||||
Qy 1166 AACAACTGTCGGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 1225
Db |||||
Qy 181 ATGGGTACCGGTGTTGGTATCCCGTCTGCTCACTACACCAAGAACTGATCACCAGT 240
Db |||||
Qy 1226 ATGGGTACCGGTATGGGTATCCCGTCTGCTCACTACACCAAGAACTGATCACCAGT 1285
Db |||||
Qy 241 TTGGCGCTGAAGAAATTTATCCGCGTGGGTTCCTGTGGCGAGTTCGCCGCACTGAAA 300
Db |||||
Qy 1286 TTGGCGCTGAAGAAATTTATCCGCGTGGGTTCCTGTGGCGAGTTCGCCGCACTGAAA 1345
Db |||||
Qy 301 CTGGCGACCTGCTTATCGGTATGGGTGCTGCAACCGATTCCAAAGTTAACCGCATCCGT 360
Db |||||
Qy 1346 CTGGCGACCTGCTTATCGGTATGGGTGCTGCAACCGATTCCAAAGTTAACCGCATCCGT 1405
Db |||||
Qy 361 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGCCTGAAAGCGCAGTAGAT 420
Db |||||
Qy 1406 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGCCTGAAAGCGCAGTAGAT 1465
Db |||||
Qy 421 GCAGCTAAAGCACATGGGTATTTGATGCTGGGTGGGTAACTGTTCTCGCGTGAACCTGTTTC 480
Db |||||
Qy 1466 GCAGCTAAAGCACATGGGTATTTGATGCTGGGTGGGTAACTGTTCTCGCGTGAACCTGTTTC 1525
Db |||||
Qy 481 TACTCTCCGACGCGGAAATGTTCCGACTGATGAAATAACGGCATTCCTCGCGTGGAA 540
Db |||||
Qy 1526 TACTCTCCGACGCGGAAATGTTCCGACTGATGAAATAACGGCATTCCTCGCGTGGAA 1585
Db |||||
Qy 541 ATGAAAGCGGTGCTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Db |||||
Qy 1586 ATGAAAGCGGTGCTATCTACGGGTGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 1645
Db |||||
Qy 601 TGCACCGTATCTGACCATCCGCACTCAGAGCAGACCACTGCGCGTGAAGCTGAGCTCAGACT 660
Db |||||
Qy 1646 TGCACCGTATCTGACCATCCGCACTCAGAGCAGACCACTGCGCGTGAAGCTGAGCTCAGACT 1705
Db |||||
Qy 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 720
Db |||||
Qy 1706 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAGTAA 1765
Db |||||

RESULT 12

US-11-036-497-5
; Sequence 5, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; FILE REFERENCE: 02901/000J410-USO
; CURRENT APPLICATION NUMBER: US/11/036,497
; PRIOR FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 4189


```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid
; FEATURE:
; NAME/KEY: gene
; LOCATION: (243)..(1021)
; OTHER INFORMATION: udp
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1037)..(1766)
; OTHER INFORMATION: deod
US-11-036-497-5
```

```
Query Match          99.6%; Score 716.8; DB 24; Length 4189;
Best Local Similarity 99.7%; Pred. No. 6.7e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAAATGCAGAAATGGCGATTTCGCTGACGTAGTATTTGATGCCA 60
DB 1046 ATGGCTACCCACACATTAAATGCAGAAATGGCGATTTCGCTGACGTAGTATTTGATGCCA 1105

QY 61 GGGACCCGCTGCGTGCAGAGTATATTCGTGAACATTTCTTGAAGATGCCCGTGAAGTG 120
DB 1106 GGGACCCGCTGCGTGCAGAGTATATTCGTGAACATTTCTTGAAGATGCCCGTGAAGTG 1165

QY 121 AACAAAGTTTCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
DB 1166 AACAAAGTTTCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 1225

QY 181 ATGGGTCAACGGTGTGGTATCCCGTCTCGTCCATCTACACCAAGAACTGATCACCGAT 240
DB 1226 ATGGGTCAACGGTATGGGTATCCCGTCTCGTCCATCTACACCAAGAACTGATCACCGAT 1285

QY 241 TTCCGGCGTGAAGAAAATTAATCCGGTGGGTTCCTGTGGCGAGTTCTGCGCGACGTAAAA 300
DB 1286 TTCCGGCGTGAAGAAAATTAATCCGGTGGGTTCCTGTGGCGAGTTCTGCGCGACGTAAAA 1345

QY 301 CTGGGACGCTGTTATCGGTATGGGTGCTGCACCGATTCCTCAAGTTTACCCGATCCGT 360
DB 1346 CTGGGACGCTGTTATCGGTATGGGTGCTGCACCGATTCCTCAAGTTTACCCGATCCGT 1405

QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGGTAAACGGATAGAT 420
DB 1406 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGGTAAACGGATAGAT 1465

QY 421 GCAGCTAAAGCACTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCGCTGACCTGTTTC 480
DB 1466 GCAGCTAAAGCACTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCGCTGACCTGTTTC 1525

QY 481 TACTCTCCGACCGCGAAATGTTCCGAGTGTGGAATAACGGATTTCTCGCGGTGGAA 540
DB 1526 TACTCTCCGACCGCGAAATGTTCCGAGTGTGGAATAACGGATTTCTCGCGGTGGAA 1585

QY 541 ATGGAAGCGCTGGTATCTACGGCGTTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
DB 1586 ATGGAAGCGCTGGTATCTACGGCGTTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 1645

QY 601 TGCAACCGTATCTGACCAATCTCGCACTCAAGAGACACCACTGCGCGTGAAGCGTCAAGT 660
DB 1646 TGCAACCGTATCTGACCAATCTCGCACTCAAGAGACACCACTGCGCGTGAAGCGTCAAGT 1705

QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATTCGTTCTGCTGGCGGATTAAGAGTAA 720
DB 1706 ACCTTCAACGACATGATCAAAATCGCACTGGAATTCGTTCTGCTGGCGGATTAAGAGTAA 1765
```

```
RESULT 13
US-10-198-034-5
; Sequence 5, Application US/10198034
; Publication No. US20030077268A1
; GENERAL INFORMATION:
; APPLICANT: Sorscher, Eric J.
```

```
; APPLICANT: Bennett, Leonard L.
; APPLICANT: Parker, William B.
; APPLICANT: Waud, William
; APPLICANT: Gadi, Vijaykrishna K.
; TITLE OF INVENTION: RECOMBINANT BACTERIAL CELLS FOR DELIVERY OF PNP TO TUMOR CELLS
; FILE REFERENCE: UAB-12407/22
; CURRENT APPLICATION NUMBER: US/10/198,034
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 08/702,181
; PRIOR FILING DATE: 1996-08-23
; PRIOR APPLICATION NUMBER: 08/122,321
; PRIOR FILING DATE: 1993-09-14
; PRIOR APPLICATION NUMBER: 60/064,676
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 5013
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-198-034-5
```

```
Query Match          99.6%; Score 716.8; DB 14; Length 5013;
Best Local Similarity 99.7%; Pred. No. 7.2e-228;
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAAATGCAGAAATGGCGATTTCGCTGACGTAGTATTTGATGCCA 60
DB 413 ATGGCTACCCACACATTAAATGCAGAAATGGCGATTTCGCTGACGTAGTATTTGATGCCA 472

QY 61 GGGACCCGCTGCGTGCAGAGTATATTCGTGAACATTTCTTGAAGATGCCCGTGAAGTG 120
DB 473 GGGACCCGCTGCGTGCAGAGTATATTCGTGAACATTTCTTGAAGATGCCCGTGAAGTG 532

QY 121 AACAAAGTTTCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
DB 533 AACAAAGTTTCGGGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 592

QY 181 ATGGGTCAACGGTGTGGTATCCCGTCTCGTCCATCTACACCAAGAACTGATCACCGAT 240
DB 593 ATGGGTCAACGGTATGGGTATCCCGTCTCGTCCATCTACACCAAGAACTGATCACCGAT 652

QY 241 TTCCGGCGTGAAGAAAATTAATCCGGTGGGTTCCTGTGGCGAGTTCTGCGCGACGTAAAA 300
DB 653 TTCCGGCGTGAAGAAAATTAATCCGGTGGGTTCCTGTGGCGAGTTCTGCGCGACGTAAAA 712

QY 301 CTGGGACGCTGTTATCGGTATGGGTGCTGCACCGATTCCTCAAGTTTACCCGATCCGT 360
DB 713 CTGGGACGCTGTTATCGGTATGGGTGCTGCACCGATTCCTCAAGTTTACCCGATCCGT 772

QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGGTAAACGGATAGAT 420
DB 773 TTTAAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTGGGTAAACGGATAGAT 832

QY 421 GCAGCTAAAGCACTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCGCTGACCTGTTTC 480
DB 833 GCAGCTAAAGCACTGGGTATGATGCTCGCGTGGGTAAACCTGTTCTCGCTGACCTGTTTC 892

QY 481 TACTCTCCGACCGCGAAATGTTTCGAGTGTGGAATAACGGATTTCTCGCGGTGGAA 540
DB 893 TACTCTCCGACCGCGAAATGTTTCGAGTGTGGAATAACGGATTTCTCGCGGTGGAA 952

QY 541 ATGGAAGCGCTGGTATCTACGGCGTTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
DB 953 ATGGAAGCGCTGGTATCTACGGCGTTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 1012

QY 601 TGCAACCGTATCTGACCAATCTCGCACTCAAGAGACACCACTGCGCGTGAAGCGTCAAGT 660
DB 1013 TGCAACCGTATCTGACCAATCTCGCACTCAAGAGACACCACTGCGCGTGAAGCGTCAAGT 1072

QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATTCGTTCTGCTGGCGGATTAAGAGTAA 720
DB 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATTCGTTCTGCTGGCGGATTAAGAGTAA 1132
```


Qy	541	ATGGAAGCGCTGGTATCTACGGGTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC	600
Db	953	ATGGAAGCGGCTGGTATCTACGGGTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC	1012
Qy	601	TGCACCGTATCTGCACACATCCGCACTCACGAGCAGACCACTGCGCGTGAGCGTCAGACT	660
Db	1013	TGCACCGTATCTGCACACATCCGCACTCACGAGCAGACCACTGCGCGTGAGCGTCAGACT	1072
Qy	661	ACCTTCACGACATGATCAAAATCGCACTCGGAATCCGTTCTGCTGGGCGGATAAAGAGTAA	720
Db	1073	ACCTTCACGACATGATCAAAATCGCACTCGGAATCCGTTCTGCTGGGCGGATAAAGAGTAA	1132
RESULT 15			
US-09-891-865A-7			
; Sequence 7, Application US/09891865A			
; Publication No. US20030059870A1			
; GENERAL INFORMATION:			
; APPLICANT: NORPHARMA SPA			
; TITLE OF INVENTION: Recombinant bacterial strains for the production of			
; FILE REFERENCE: 99DC26E			
; CURRENT APPLICATION NUMBER: US/09/891,865A			
; CURRENT FILING DATE: 2002-06-25			
; PRIOR APPLICATION NUMBER: MI98A002792			
; PRIOR FILING DATE: 1998-12-23			
; NUMBER OF SEQ ID NOS: 15			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 7			
; LENGTH: 5241			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid			
; NAME/KEY: gene			
; LOCATION: (1312)..(2042)			
; OTHER INFORMATION: deod			
US-09-891-865A-7			
Query Match 99.6%; Score 716.8; DB 10; Length 5241;			
Best Local Similarity 99.7%; Pred. No. 7.4e-228;			
Matches 718; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	1	ATGGCTACCCACACACATTAATGCAGAAATGGGGGATTTTCGCTGACGTAGTTTTGATGCCA	60
Db	1322	ATGGCTACCCACACACATTAATGCAGAAATGGGGGATTTTCGCTGACGTAGTTTTGATGCCA	1381
Qy	61	GGCGACCGCTGGTGGGAAGTATATTGCTGAAACTTTTCTTGAAGATGCCGCTGAAGTG	120
Db	1382	GGCGACCGCTGGTGGGAAGTATATTGCTGAAACTTTTCTTGAAGATGCCGCTGAAGTG	1441
Qy	121	AACAAAGTTTCGCGGTATGCTGGGCTTCCACGGTACTTTACAAAGCCGCCAAATTTCCGTA	180
Db	1442	AACAAAGTTTCGCGGTATGCTGGGCTTCCACGGTACTTTACAAAGCCGCCAAATTTCCGTA	1501
Qy	181	ATGGGTACACGGTGTGGTATCCGCTCCTGCTCCATCTACACCAAGAACTGATCACCCGAT	240
Db	1502	ATGGGTACACGGTATGGTATCCGCTCCTGCTCCATCTACACCAAGAACTGATCACCCGAT	1561
Qy	241	TTGGCGGTGAAGAAATATTATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGCAGTAAAA	300
Db	1562	TTGGCGGTGAAGAAATATTATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGCAGTAAAA	1621
Qy	301	CTGCGGACGCTCGTTATCCGTTATGGGTGCTGCAACCGATTCCAAAGTTAAACCGCATCCGT	360
Db	1622	CTGCGGACGCTCGTTATCCGTTATGGGTGCTGCAACCGATTCCAAAGTTAAACCGCATCCGT	1681
Qy	361	TTTAAAGACCATGACTTTTGGCGTATCGCTGACTTCGACATGTTGTCGTAAACGCAAGTAGAT	420
Db	1682	TTTAAAGACCATGACTTTTGGCGTATCGCTGACTTCGACATGTTGTCGTAAACGCAAGTAGAT	1741

Qy	421	GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTT	480
Db	1742	GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTT	1801
Qy	481	TACTCTCCGACCGCGGAAATGTTTCGACGTGATGGAAAAATACGGCATTTCTCGGCGTGGAA	540
Db	1802	TACTCTCCGACCGCGGAAATGTTTCGACGTGATGGAAAAATACGGCATTTCTCGGCGTGGAA	1861
Qy	541	ATGGAAAGCGGCTGGTATCTACGGGTGCTGCGAGNATTTGGCGGAAAGCCCTGACCATC	600
Db	1862	ATGGAAAGCGGCTGGTATCTACGGGTGCTGCGAGNATTTGGCGGAAAGCCCTGACCATC	1921
Qy	601	TGCACCGTATCTGACCAACATCCGCACCTCACGAGCAGACCACTGCCGCTGAGCGTCAGACT	660
Db	1922	TGCACCGTATCTGACCAACATCCGCACCTCACGAGCAGACCACTGCCGCTGAGCGTCAGACT	1981
Qy	661	ACCTTCAACGACATGATCAAAATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGCAT	720
Db	1982	ACCTTCAACGACATGATCAAAATCGCATCGCATCGCATCGCATCGCATCGCATCGCATCGCAT	2041

Search completed: July 23, 2005, 01:38:14
Job time : 806 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 22, 2005, 17:40:37 ; Search time 4167 Seconds
(without alignments)
6576.980 Million cell updates/sec

Title: US-10-035-300A-1

Perfect score: 720

Sequence: 1 atggctaccacacattaa.....tgctggcgataaagagtaa 720

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_ges1:*

9: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	272.6	37.9	767	7	CV205284
C 2	272.6	37.9	769	7	CV204676
C 3	272.6	37.9	771	7	CV219418
C 4	272.6	37.9	792	7	CV204675
5	272.6	37.9	794	7	CV205285
6	272.6	37.9	794	7	CV219419
7	260.4	36.2	721	7	CV20577885
C 8	191.6	26.6	1589	8	BH770690
C 9	169	23.5	169	7	CF306721
10	169	23.5	169	7	CF306860
11	150	20.8	161	7	CF307418
12	121.8	16.9	921	8	BH138974
C 13	110	15.3	863	8	AZ680406
C 14	101.8	14.1	861	8	BH166036
C 15	101	14.0	864	8	AZ675191
C 16	100	13.9	878	8	AZ533877
C 17	93.8	13.0	865	8	AZ669801
C 18	80.2	11.1	226	7	CV213810
C 19	80.2	11.1	250	7	CV213811
C 20	77	10.7	882	8	AZ681446
C 21	74	10.3	879	8	AZ539664
C 22	66	9.2	861	8	AZ669012
C 23	65	9.0	715	9	AG266877
24	64.6	9.0	234	7	CV203815

C 25	64.6	9.0	234	7	CV203816	CV203816	EST863526
C 26	59.4	8.2	904	8	AZ536366	AZ536366	ENTCX45TF
C 27	57	7.9	801	9	CC924665	t078f14ba	CC924665
C 28	56.4	7.8	898	8	BH148646	ENTPE80TR	BH148646
C 29	46.8	6.5	890	8	AZ529376	ENTCV88TF	AZ529376
30	43	6.0	742	4	BM170790	EST573313	BM170790
31	42	5.8	572	4	BM165171	EST567694	BM165171
32	41.4	5.8	602	7	CN896963	010630AAZ	CN896963
33	41.4	5.8	645	4	BM170428	EST572951	BM170428
34	41.4	5.8	690	4	BM162943	EST565466	BM162943
35	41.4	5.8	709	4	BM160915	EST563438	BM160915
36	41.4	5.8	734	4	BM159186	EST561709	BM159186
37	41.4	5.8	768	4	BM159582	EST562105	BM159582
38	40	5.6	706	6	C96538	EST562105	C96538
39	39.8	5.5	260	7	CO866368	C96538 Marc	C96538
40	39.8	5.5	543	7	CN918728	Mdbb5007g	CN918728
41	39.8	5.5	559	7	CN903541	0320707ABP	CN903541
42	39.8	5.5	604	4	BM162201	021018ABC	BM162201
43	39.8	5.5	649	7	CN902206	EST564724	CN902206
44	39.8	5.5	681	7	CN903108	021014ABC	CN903108
45	39.8	5.5	743	7	CO866081	021016ABC	CO866081

ALIGNMENTS

RESULT 1
CV205284/4/c
LOCUS CV205284 767 bp mRNA linear EST 16-SEP-2004
DEFINITION EST864994 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone rVTAC60 3' end, mRNA sequence.
ACCESSION CV205284
VERSION CV205284.1 GI:52152264
KEYWORDS EST.
SOURCE Trichomonas vaginalis
ORGANISM Trichomonas vaginalis
Eukaryota; Parabasalidea; Trichomonadida; Trichomonadidae; Trichomonadinae; Trichomonas.

REFERENCE 1 (bases 1 to 767)
AUTHORS Carlton,J.M., Dyall,S., Johnson,P.J. and Fraser,C.M.
TITLE The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis

JOURNAL Unpublished (2004)
COMMENT Other ESTs: EST864995
Contact: Jane Carlton

Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seq primer: polydT 24 base.
Location/Qualifiers

source
1..767
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TVTAC60"
/clone_lib="non-normalized T1 cDNA library"
/note="Vector: Lambda Triplex2; Site_1: SfiI; Site_2: SfiI; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiI and SfiI sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

ORIGIN

Query Match 37.9%; Score 272.6; DB 7; Length 767;
Best Local Similarity 62.0%; Pred. No. 3e-73;
Matches 431; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

```
Qy 1 ATGCTACCCACACATTAATGACAGAAATGGCGGATTTGGCTGACGTAGTTTTCATGCCA 60
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 755 ATGCAACACCCCACTACTGCAAGGTTGGCGATTTGGCTGACAGATTCATCTATGTC 696
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 GCGACCCGCTGCGTGGGAAGTATATTGCTGAAATCTTTCCTTGAAGATGCCCGTGAAGTG 120
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 695 GCGATCCACTCCGTCGCAAGCTCATTTGCTGATAACTATCTTGAGAACGCTTAAGCAAGTC 636
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 AACAACTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCCCGCAAAATTTCCGTA 180
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 635 AACTCTGTTCTGGAATGCTTGGTTTACAGGSCACATACAGGCGCAACACCTCTCTGTT 576
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 ATGGGTACAGGTGTTGGTATCCCGCTGCTGCTCACTACACCAAGAACTGATCACCGAT 240
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 575 ATGGGTATGCGATGGGTATCCCATCAATTTCCATCTACGCTGAAGAACTCTACAGCTT 516
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 241 TTCGGCGTGAAGAAATATATCCGCGTGGGTTCTGTGGCGGAGTTCTGCCGACAGTAAAA 300
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 515 TACAAGGTTAAGACCATCATCCGTTGCGCACTTGGCGCACAGTTGATCCAAATGTTTAC 456
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 CTGGCGAGCTGCTTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 360
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 GTCCGCGATGTCGATGTCACAGCATCTGGCCACAGATTCAAATGTTACAGAAATCCG 396
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 361 TTTAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGTCGTCGTAACGCGATAGAT 420
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 395 CTTCTTGGCCACGATTTCCAGCTTACAGCCAACTTCGAAGTCGTTAGTGCCTTAGTTGAA 336
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 421 GCAGCTAAAGCACTGGGTATTGATGCTCGGGTGGGTAACTGTTCTCCGCTGACCTGTT 480
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 335 TCCGCCAAGGCTCTCAACATCCCAACACAGGTTGGAAAGGCTTACTCAACAGATATCTTC 276
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 481 TACTCTCGGACGCGGAATGTTTCAGAGTGATGGAATAATACGCACTTCTCGGCTGGAA 540
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275 TACAGCAAGGAACAAGGATTAATATGAGGCGCTTCTCAGTACCACTTCATGCTGTCGA 216
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 541 ATGGAAGCGGCTGCTATCTACGGGTCGCTGACAAATTTGGCGGAAAGCCCTGACCATC 600
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 215 ATGGAATCTGCTGCTGCTTCCCAATTTGCTGATTAATGCGGCAAGAGCTGGCTGATC 156
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 601 TGCACCGTATCTGACCATCCGCACTCAGAGCAGACCACTGCGCTGAGCGTCAAGT 660
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 155 TGCACAGTTTCGGATCATCATCACACAGTCCGCTACACAGAGCGGTCACAGAGCGCAGACA 96
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 661 ACCTTCAACGATGATCAAAATCGCACTGGAATC 695
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 95 TCATTCCAGAACATGATCAAGATCGCTCTTGAAGC 61
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 2

CV204676/c
LOCUS
DEFINITION
EST864386 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone TVFA905 5' end, mRNA sequence.

ACCESSION
CV204676

VERSION
EST

KEYWORDS
CV204676.1 GI:52151656

SOURCE
Trichomonas vaginalis

ORGANISM
Trichomonas vaginalis

REFERENCE
Bukaryota; Parabasalidea; Trichomonada; Trichomonadida;

AUTHORS
Trichomonadidae; Trichomonadinae; Trichomonas.

TITLE
1 (bases 1 to 769)

The complete genome sequence of the sexually transmitted parasite

Trichomonas vaginalis

Unpublished (2004)

Other ESTs: EST864385

Contact: Jane Carlton

Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@igr.org
Seq primer: lambda Triplex2.
FEATURES
Source
1. 769
Location/Qualifiers
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TVFA905"
/clone_lib="non-normalized T1 cDNA library"
/notes="Vector: Lambda Triplex2; Site 1: SfiA; Site 2:
SfiB; T. vaginalis strain T1 library constructed from
cDNA, made in lambda Triplex2. Inserts cloned
unidirectionally in the SfiA and SfiB sites. Mass excision
of library produced inserts in pTriplex2 plasmid. Inserts
sequenced from both 5' and 3' ends using triplex2
sequencing primer and polydT 24 bp primer respectively."

ORIGIN

Query Match 37.9%; Score 272.6; DB 7; Length 769;
Best Local Similarity 62.0%; Pred. No. 3e-73;
Matches 431; Conservative 0; Mismatches 264; Indels 0; Gaps 0;
Qy 1 ATGCTACCCACACATTAATGACAGAAATGGCGGATTTGGCTGACGTAGTTTTCATGCCA 60
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 757 ATGCAACACCCCACTACTGCAAGGTTGGCGATTTGGCTGACAGATTCATCTATGTC 698
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 GCGACCCGCTGCGTGGGAAGTATATTGCTGAAATCTTTCCTTGAAGATGCCCGTGAAGTG 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 697 GCGATCCACTCCGTCGCAAGCTCATTTGCTGATAACTATCTTGAGAACGCTTAAGCAAGTC 638
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 AACAACTTCCGCGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 637 AACTCTGTTCTGGAATGCTAGGTTTACAGGACATACAAAGGCGCAACCACTCTCTGTT 578
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 ATGSGTCAACCGTGTGGTATCCCGTCTCTGCTCCATCTTACACCAAGAACTGATCACCGAT 240
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 577 ATGGTCATGCGATGGTATCCCATCACTTCCATCTAGCTGAAGAACTCTACAAGCTT 518
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 241 TTGGGCTGAAGAAATTTATCCGCGTGGGTTCTCTGCGGCGAGTTCGCGCGACGTAAAA 300
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 517 TACAAGGTTAAGACCATCATCCGTCGCGCACTTGGCGCACAGTTGATCCAAATGTTTCA 458
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 CTGGCGACCTGTTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 360
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 457 GTCCGCGATGCTGCAATGTCACAGATCTGGGCACAGATTCAAATGTTACAGAAATGCC 398
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 361 TTTAAGACCATGACTTTGCGGCTATCGGTGACTTCGACATGGTGGGTAAACGAGTAGAT 420
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 397 CTTCTTGGCCACGATTTCCAGCTTACAGCCAACTTCGAAGTCGTTAGTGCCTTAGTTGA 338
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGGCTGGGTAACTGTTCTCCGCTGACCTGTT 480
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 337 TCGCGCAAGGCTCTCAACATCCCAACACAGGTTGGAAAGGCGCTACTCAACAGATATCTTC 278
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 481 TACTCTCCGACGCGGAAATGTTTCGAGCTGATCGAAAAATACGCGCATTCGCGGCTGGAA 540
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 277 TACAGCAAGGACAGGATTAATGAGGCCCTTGTCTCAGTACCACCTTCATTGCTGTGCA 218
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 541 ATGGAAGCGGCTGGTATCTACGGCGTGGTTCGAGAAATTTGGCGCGAAAGCCCTGACCATC 600
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 217 ATGGAATCTGCTGGTCCCTTCCCAATTTGCTGATTACTATATGCGCAAGAGCTGGCTGCATC 158
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 601 TGACCGTATCTACACCATCCGCACTCAGGACGACACCACTGCGCTGAGCGTCAAGT 660
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 157 TGCACAGTTTTCGGATCACATCATCACAGAGTTCGCTACACCGGAAGAGCGCCAGACA 98
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 661 ACCTTCAACGATGATCAAAATCGCACTGGAATC 695
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 97 TCATTCCAGAACATGATCAAGATCGCTCTTGAAGC 63
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3

```

CV219418/c
LOCUS      CV219418      771 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION EST879128 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone TVTCX26 3' end, mRNA sequence.
ACCESSION  CV219418
VERSION     CV219418.1
KEYWORDS   CV219418.1 GI:52166398
SOURCE     EST.
ORGANISM   Trichomonas vaginalis
            Trichomonas vaginalis
            Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
            Trichomonadidae; Trichomonadinae; Trichomonas.
            1 (bases 1 to 771)
            Carlton, J.M., Dyall, S., Johnson, P.J. and Fraser, C.M.
            The complete genome sequence of the sexually transmitted parasite
            Trichomonas vaginalis
            Unpublished (2004)
            Other ESTs: EST879129
            Contact: Jane Carlton
            Parasite Genomics Group
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-530-9319
            Fax: 301-838-0208
            Email: carlton@tigr.org
            Seq primer: polydT 24 base.
            Location/Qualifiers
                1..771
                /organism="Trichomonas vaginalis"
                /mol_type="mRNA"
                /strain="T1"
                /db_xref="taxon:5722"
                /clone="TVTCX26"
                /note="Vector: Lambda Triplex2; Site 1: SfiA; Site 2:
                SfiB; T. vaginalis strain T1 library constructed from
                cDNA, made in lambda Triplex2. Inserts cloned
                unidirectionally in the SfiA and SfiB sites. Mass excision
                of library produced inserts in pTriplex2 plasmid. Inserts
                sequenced from both 5' and 3' ends using Triplex2
                sequencing primer and polydT 24 bp primer respectively."

FEATURES             source
    Query Match      37.9%; Score 272.6; DB 7; Length 771;
    Best Local Similarity 62.0%; Pred. No. 3e-73;
    Matches 431; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCGAAGTATATTGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 60
DB 759 ATGGCAACACCCCATTAATCTGCAAAAGTTGGCGATTTGCGTGAGACAGTTCTCATGTGC 700

QY 61 GCGCACCCGCTGCGTGCAGAGTATATTGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 120
DB 699 GCGGATCCACTCCGTCGCCAGCTCATTTGCTGTAATCTTGAAGAGCTTAAGCAAGTC 640

QY 121 AACACAGTTTCGCGGTATGCTGGGCTTCAACGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
DB 639 AACTCTGTTTCGTAATGCTTGGTTTACAGGCACATACAAAGGCAAAACCACTCTCTGTT 580

QY 181 ATGGGTACAGGTGTTGGTATCCCGTCTGCTCCATCTACCAAGAACTGATCAGCAT 240
DB 579 ATGGGTATGGCATGGGTATCCCATTTCCATCTACGCTGAAGAACTCTACAAAGCTT 520

QY 241 TTCGCGGTGAAGAAATTTATCCGCTGGGTTCCTGTGGCGAGTTCTCCCGCACGTAAGA 300
DB 519 TACAAGGTTAAGACCATCATCCGTTGCGCACTTGGCGCACAGTTGATCCAAATGTTTAC 460

QY 301 CTGGCGGACGTCGTTATCGGTATGGGTGCGTGCACCGGATTCAAAGTTTAAACCGCATCCGT 360
DB 459 GTCCGCGCATGTCTGCATTTGTACAGCATCTGGCACAGATTCAAATGTTTAAACAGATGGC 400

QY 361 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTTCGTAACGCGATGAT 420
DB 399 CTTCCTGGCCACGATTTTCCCGAGCTACAGCCAACTTCGAAGTCGTTAGTGCCTTAGTTGAA 340

```

```

CV204675
LOCUS      CV204675      792 bp      mRNA      linear      EST 16-SEP-2004
DEFINITION EST864385 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone TVTA905 3' end, mRNA sequence.
ACCESSION  CV204675
VERSION     CV204675.1
KEYWORDS   CV204675.1 GI:52151655
SOURCE     EST.
ORGANISM   Trichomonas vaginalis
            Trichomonas vaginalis
            Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
            Trichomonadidae; Trichomonadinae; Trichomonas.
            1 (bases 1 to 792)
            Carlton, J.M., Dyall, S., Johnson, P.J. and Fraser, C.M.
            The complete genome sequence of the sexually transmitted parasite
            Trichomonas vaginalis
            Unpublished (2004)
            Other ESTs: EST864386
            Contact: Jane Carlton
            Parasite Genomics Group
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-530-9319
            Fax: 301-838-0208
            Email: carlton@tigr.org
            Seq primer: polydT 24 base.
            Location/Qualifiers
                1..792
                /organism="Trichomonas vaginalis"
                /mol_type="mRNA"
                /strain="T1"
                /db_xref="taxon:5722"
                /clone="TVTA905"
                /note="Vector: Lambda Triplex2; Site 1: SfiA; Site 2:
                SfiB; T. vaginalis strain T1 library constructed from
                cDNA, made in lambda Triplex2. Inserts cloned
                unidirectionally in the SfiA and SfiB sites. Mass excision
                of library produced inserts in pTriplex2 plasmid. Inserts
                sequenced from both 5' and 3' ends using Triplex2
                sequencing primer and polydT 24 bp primer respectively."

ORIGIN

    Query Match      37.9%; Score 272.6; DB 7; Length 792;
    Best Local Similarity 63.0%; Pred. No. 3.1e-73;
    Matches 431; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 1 ATGGCTACCCACACATTAATGCGAAGTATGCGGATTTGCGTACGTTAGTTTTCATGCCA 60
DB 13 ATGGCAACACCCCATTAATCTGCAAAAGTTGGCGATTTGCGTGAGACAGTTCTCATGTGC 72

```

```
QY 61 GGCACCGCGTGGTGGAGTATATGCTGAATCTTTCCTTGAAGATGCCGTTGAAGTG 120
D 61 |||||
QY 73 GGCATCACTCCGTCGCAAGCTCATTTGCTGATCACTATCTTGAGACGCTAAGCAAGTC 132
D 73 |||||
QY 121 AACAACTTTCCGGTATGCTGGGCTTCCACCGTACTTACAAAGCGCCGCAAAATTTCCGTA 180
D 121 |||||
QY 133 AACTCTGTTGCTGGAATGCTAGGTTTACAGGCACATACAAAGGCAACACCTCTCTGTT 192
D 133 |||||
QY 181 ATGGGTACGGTGTGGTATCCCGTCTCTGCTCATCTACACCAAGAACTGATCAGCGAT 240
D 181 |||||
QY 193 ATGGGTATGGCATGGGTATCCCATCTCCATCTACGCTGAGAACTCTACAAGCTT 252
D 193 |||||
QY 241 TTCCGGCTGAAGAAATATTCACCGGTGGGTTCTGCTGGCGCAGTTCTGCGCAGTAA 300
D 241 |||||
QY 253 TACAAAGTTAAGACCATCATCCGTTGCGGACATTTGCGGACAGTTGATCCAAATGTTCC 312
D 253 |||||
QY 301 CTGCGGACGTCGTTATCGGTATGGGTGCTGCAACCGATTCCTCAAGTTAACCGCATCCGT 360
D 301 |||||
QY 313 GTCCGCGATGTCGATTTGTCACAGCATCTGGCACAGATTCCTCAAGTTTAAACAGAAATGCGC 372
D 313 |||||
QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGATCGCTGATCGCTGATCGCTGATCGCTGAT 420
D 361 |||||
QY 373 CTTCTTGGCGACGATTTCCGAGCTACAGCACTTTCGAGTCTGTTAGTCTAGTTGAA 432
D 373 |||||
QY 421 GCAGCTAAAGCAGTGGGTATTTGATGCTCGGCTGAGTAACTGTTCTCGCTGACCTGTT 480
D 421 |||||
QY 433 TCCGCAAGGCTCTCAACATCCCAACAGAGTTGGAAGGCGCTACTCAACAGATATCTTC 492
D 433 |||||
QY 481 TACTCTCCGACGCGGAAATTTTCGACGTGATGGAATAATACGCAATTTCTCGCGTGGAA 540
D 481 |||||
QY 493 TACAGATTTCCGATCATCATCATCACACAGATTCGCGTACACGGAAGAGCGCCAGACA 552
D 493 |||||
QY 541 ATGGAAGCGGCTGATCTACGGGCTGCTGACAGAAATTTGCGGGAAGGCGCTGACCATC 600
D 541 |||||
QY 553 ATGGAATCTGCTGGTCCCTTCCCAATTTGCTGATTAATGCTGAGTCTGAGTCTGATC 612
D 553 |||||
QY 601 TGACCCGCTATCTGACCATCTCGGCTACAGAGCAGACCACTGCGGCTGAGCGTCAAGT 660
D 601 |||||
QY 613 TGCACATTTCCGATCATCATCATCACACAGATTCGCGTACACGGAAGAGCGCCAGACA 672
D 613 |||||
```

RESULT 5

```
CV205285 794 bp mRNA linear EST 16-SEP-2004
LOCUS EST864995 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
DEFINITION clone TVTAC60 5' end, mRNA sequence.
```

```
CV205285
VERSION CV205285.1 GI:52152265
```

KEYWORDS

```
EST.
```

ORGANISM

```
Trichomonas vaginalis
Trichomonas vaginalis
Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
Trichomonadidae; Trichomonadinae; Trichomonas.
```

REFERENCE

```
1 (bases 1 to 794)
Carlton,J.M., Dyall,S., Johnson,P.J. and Fraser,C.M.
The complete genome sequence of the sexually transmitted parasite
Trichomonas vaginalis
Unpublished (2004)
```

JOURNAL

```
Other ESTs: EST864994
```

COMMENT

```
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@igr.org
```

FEATURES

```
Seq primer: lambda Triplex2.
```

source

```
Location/Qualifiers
```

```
1..794
```

```
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
```

```
/clone="TVTAC60"
```

```
/notes="Vector: Lambda Triplex2; Site 1: SfiA; Site 2:
```

```
SfiB; T. vaginalis strain T1 library constructed from
```

```
cDNA, made in lambda triplex2. Inserts cloned
```

```
unidirectionally in the SfiA and SfiB sites. Mass excision
```

```
of library produced inserts in pTriplex2 plasmid. Inserts
```

```
sequenced from both 5' and 3' ends using Triplex2
```

```
sequencing primer and polydT 24 bp primer respectively."
```

ORIGIN

```
Query Match 37.9%; Score 272.6; DB 7; Length 794;
Best Local Similarity 62.0%; Pred. No. 3.1e-73;
Matches 431; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 1 ATGCTACCCACACATTAATGCAAAATGGGCGATTTGCTGACGTAGTTTGTATGCCA 60
D 1 |||||
QY 13 ATGGCAACACCCCATAACTCTGCAAGGTTGGCGATTTGCTGAGACAGTTCTCATGTGC 72
D 13 |||||
QY 61 GGGACCCCGCTGGTGGAGTATATTTGCTGAAACTTTCTTGAAGATGCCGTTGAAGTG 120
D 61 |||||
QY 73 GGGATCCACTCGTGGCAAGCTCATTTGCTGATAACTATCTTGAGAAGCTTAAGCAAGTC 132
D 73 |||||
QY 121 AACAACTTTCCGGTATGCTGGGCTTACCGGTACTTACAAAGCGCCGCAAAATTTCCGTA 180
D 121 |||||
QY 133 AACTCTGTTGCTGGAATGCTTGGTTTTACAGGCACATACAAAGGCGCAACACCTCTCTGTT 192
D 133 |||||
QY 181 ATGGGTACGGTGTGGTATCCCGTCTGCTGATCGCTGATCGCTGATCGCTGATCGCTGAT 240
D 181 |||||
QY 193 ATGGGTATGGCATGGGTATCCCATCTCCATCTACGCTGAGAACTCTACAAGCTT 252
D 193 |||||
QY 241 TTCCGGCTGAAGAAATATTCGCGGTGGGTTCTGCTGGCGCAGTTCTGCGCAGTAA 300
D 241 |||||
QY 253 TACAAAGTTAAGACCATCATCCGTTGCGGACATTTGCGGACAGTTGATCCAAATGTTCC 312
D 253 |||||
QY 301 CTGCGGACGTCGTTATCGGTATGGGTGCTGACAGATTCCTCAAGTTAACCGCATCCGT 360
D 301 |||||
QY 313 GTCCGCGATGTCGATTTGTCACAGCATCTGGCACAGATTCCTCAAGTTTAAACAGAAATGCGC 372
D 313 |||||
QY 361 TTTAAAGACCATGACTTTGCGGCTATCGCTGATCGCTGATCGCTGATCGCTGATCGCTGAT 420
D 361 |||||
QY 373 CTTCTTGGCGACGATTTCCGAGCTACAGCACTTTCGAGTCTGTTAGTCTAGTTGAA 432
D 373 |||||
QY 421 GCAGCTAAAGCAGTGGGTATTTGATGCTCGGCTGAGTAACTGTTCTCGCTGACCTGTT 480
D 421 |||||
QY 433 TCCGCAAGGCTCTCAACATCCCAACAGAGTTTAAATGAGGCGCTACTCAACAGATATCTTC 492
D 433 |||||
QY 481 TACTCTCCGACGCGGAAATTTTCGACGTGATGGAATAATACGCAATTTCTCGCGTGGAA 540
D 481 |||||
QY 493 TACAGATTTCCGATCATCATCATCACACAGATTCGCGTACACGGAAGAGCGCCAGACA 552
D 493 |||||
QY 541 ATGGAAGCGGCTGATCTACGGGCTGCTGACAGAAATTTGCGGGAAGGCGCTGACCATC 600
D 541 |||||
QY 553 ATGGAATCTGCTGGTCCCTTCCCAATTTGCTGATTAATGCTGAGTCTGAGTCTGATC 612
D 553 |||||
QY 601 TGACCCGCTATCTGACCATCTCGGCTACAGAGCAGACCACTGCGGCTGAGCGTCAAGT 660
D 601 |||||
QY 613 TGCACATTTCCGATCATCATCATCACACAGATTCGCGTACACGGAAGAGCGCCAGACA 672
D 613 |||||
QY 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATC 695
D 661 |||||
QY 673 TCATTCCAGAACATGATCAAGATCGCTCTTGAAGC 707
D 673 |||||
```

RESULT 6

```
CV219419
```

```
LOCUS
```

```
DEFINITION
```

```
CV219419 794 bp mRNA linear EST 16-SEP-2004
```

```
EST879129 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
```

```
clone TVTAC26 5' end, mRNA sequence.
```

```

ACCESSION   CV219419.1  GI:52166399
VERSION      EST.
KEYWORDS     Trichomonas vaginalis
SOURCE       Trichomonas vaginalis
ORGANISM     Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
              Trichomonadidae; Trichomonadinae; Trichomonas.
REFERENCE    1 (bases 1 to 794)
AUTHORS      Carlton, J.M., Dyall, S., Johnson, P.J. and Fraser, C.M.
TITLE        The complete genome sequence of the sexually transmitted parasite
              Trichomonas vaginalis
JOURNAL      Unpublished (2004)
COMMENT      Other ESTs: EST879128
              Contact: Jane Carlton
              Parasite Genomics Group
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850, USA
              Tel: 301-530-9319
              Fax: 301-838-0208
              Email: carlton@tigr.org
              Seq primer: lambda Triplex2.
              Location/Qualifiers
FEATURES             1..794
                     /organism="Trichomonas vaginalis"
                     /mol_type="mRNA"
                     /strain="T1"
                     /db_xref="taxon:5722"
                     /clone="TVTCX26"
                     /clone_lib="non-normalized T1 cDNA library"
                     /note="Vector: Lambda triplex2; Site_1: SfiI; Site_2:
                     SfiB; T. vaginalis strain T1 library constructed from
                     cDNA, made in lambda Triplex2. Inserts cloned
                     unidirectionally in the SfiI and SfiB sites. Mass excision
                     of library produced inserts in pTriplex2 plasmid. Inserts
                     sequenced from both 5' and 3' ends using Triplex2
                     sequencing primer and polydr 24 bp primer respectively."
ORIGIN
Query Match      37.9%; Score 272.6; DB 7; Length 794;
Best Local Similarity 62.0%; Pred. No. 3.1e-73;
Matches 431; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

Qy 1 ATGGCTACCCACACATTAATGCGAAATGGCGGATTCGCTGACGTAGTTTGTGATGCCA 60
Db |||||
Qy 13 ATGGCAACACCCCATAACTCTGCAAGGTTGGCGATTTTCGCTGACAGATTTCTCATGTGC 72
Db |||||
Qy 61 GGCGACCGCTGCGTGGCAAGTATATGCTGAATTTTCCTTGAAGTCCCGTGAAGTG 120
Db |||||
Qy 73 GGCGATCCACTCCGTGCCAAGCTCATCTGCTAATCTCTGAGAACGCTAAGCAAGTC 132
Db |||||
Qy 121 AACACAGTTCGCGGTATGCTGGGCTTCAACCGGTACTTACAAAGGCGCAAAATTTCCGTA 180
Db |||||
Qy 133 AACTCTGTTCTGGNAATGCTTGGTTTACAGGCACATACAGGGCAACACCACTCTCTGTT 192
Db |||||
Qy 181 ATGGGTACCGGTGTTGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGT 240
Db |||||
Qy 193 ATGGGTATGTCATGGGTATCCCATCCATCTCCATCTACGCTGAGAACTCTACAACGTT 252
Db |||||
Qy 241 TTGCGCGTGAAGAAATATCCGGTGGGTTCTGTTGGCGAGTTCTGCGCAGTAA 300
Db |||||
Qy 253 TACAAGGTTAAGACATCATCCGTGTCGGCACTTGGCGCACAGTTGATCCAAATGTTCCAC 312
Db |||||
Qy 301 CTGGCGCAGCTGCTGTTATCCGTATGGGTGCTGCACCGGATTCAAAGTTAAACCGATCCGT 360
Db |||||
Qy 313 GTCCGCGATGCTGCATGTCACAGCATCTGCGCACAGATTCAAATGTTAAGATGCGC 372
Db |||||
Qy 361 TTTAAAGACCAAGACTTTGCGCGTATCCGCTGCTGATTCGACATGGTGCCTGAAACGCAAGTAG 420
Db |||||
Qy 373 CTTCTTGCGCCACGATTTCCAGCTACAGCCAACTTCGAACTCGTAGTGCCTTAGTTGAA 432
Db |||||
Qy 421 CGAGCTAAGCACTGGGTATTGATGCTCGCTGGGTAACTGTTCTCGCTGACCTGTT 480
Db |||||
Qy 433 TCCGCCAAGGCTCTCAACATCCCAACACAGGTTGGAAAGGCGCTACTCAACAGATATCTTC 492

```

```

481 TACTCTCCGGACGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGGCGTGAA 540
493 TACAGCAAGGAACAAGGATTAATAGAGCCCTTCTCAGTACCACCTTCTATTGCTGTCGA 552
541 ATGGAAGCGGCTGTATCTACGGCGTGCCTGCAGAAATTTGGCGGAAGCCCTGACCATC 600
553 ATGGAATCTGCTGCTCCCTTTCCCAATTTGCTGATAACTATGCGCAAGAGCTGGCTGCATC 612
601 TGCACCGTATCTGACCAACATCCGCACTCACGACGAGACCACTGCGGCTGAGCGTCAGACT 660
613 TGCACAGTTTCGATCATCATCATCACACGAGTCCGCTACACAGAGAGCGCCAGACA 672
661 ACCTTCAACGACATGATCAAAATCCGACTGGAATC 695
673 TCATTCAGAACATGATCAAGATCGCTCTTGAAGC 707

C0577885 721 bp mRNA linear EST 20-JUL-2004
TVEST085809 Tv30236_PT cDNA Library Trichomonas vaginalis cDNA 5',
mRNA sequence.
C0577885 GI:50408335
EST.
SOURCE Trichomonas vaginalis
ORGANISM Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
              Trichomonadidae; Trichomonadinae; Trichomonas.
REFERENCE 1 (bases 1 to 721)
AUTHORS Zhou, Y., Shu, M., Huang, S.C.C., Huang, K.Y. and Tang, P.
TITLE Analysis of Gene Expression Profile in Trichomonas vaginalis by EST
              Sequencing
JOURNAL Unpublished (2003)
COMMENT Contact: Tang, P.
              Molecular Regulation and Bioinformatics Laboratory, College of
              Medicine
              Chang Gung University
              259 Wenhsa 1st. Road, Kweishan, Taoyuan 333, Taiwan
              Tel: +886 3 3283016 EXT5136
              Fax: +886 3 3283031
              Email: petang@mail.cgu.edu.tw
              PCR Primers
              FORWARD: T7
              BACKWARD: T3
              Seq primer: T3.
              Location/Qualifiers
FEATURES             1..721
                     /organism="Trichomonas vaginalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:5722"
                     /cell_line="ATCC30236"
                     /dev_stage="trophozoites at mid-log phase"
                     /lab_host="XLI Blue-MRF"
                     /clone_lib="TV30236_PT cDNA Library"
                     /note="Vector: Lambda ZAP-Express (Stratagene); Site_1:
                     EcoRI; Site_2: XhoI"
ORIGIN
Query Match      36.2%; Score 260.4; DB 7; Length 721;
Best Local Similarity 61.8%; Pred. No. 1.9e-69;
Matches 414; Conservative 0; Mismatches 256; Indels 0; Gaps 0;

Qy 28 ATGGGCGATTTTCGCTGACGTAGTTTTCATGCCAGGCGACCGCTCGTCCGCAAGTATATT 87
Db |||||
Qy 41 AGGGCGCATTTTCGCTGAAACAGTCTCATGTGCGGTGATCCACTCCCGGCTCAAGCTCAAT 100
Db |||||
Qy 88 GCTGAATCTTCTTGAAGATGCCGTGAAGTGAACCAACAGTTTCGCGGTATGCTCGGGCTTC 147
Db |||||
Qy 101 GCTGAGACATATCTTGAATAATCCAAAGCTTGTCAACATATGTTCTGGCATTCAGAGCTAC 160
Db |||||
Qy 148 ACCGGTACTTACAAAGGCGCAAAATTTCCGTAATGGGTACGGTGTGGTATCCCGTCC 207

```

[illegible]

RESULT 8	
BH770690/c	
LOCUS	1589 bp DNA linear GSS 01-MAY-2002
DEFINITION	LMGtag441 MG1363 Random Sequence Tag Library <i>Lactococcus lactis</i> subsp. cremoris genomic, genomic survey sequence.
ACCESSION	BH770690
VERSION	BH770690.1 GI:20373647
KEYWORDS	GSS.
SOURCE	<i>Lactococcus lactis</i> subsp. cremoris
ORGANISM	<i>Lactococcus lactis</i> subsp. cremoris
REFERENCE	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; <i>Lactococcus</i> .
AUTHORS	1 (bases 1 to 1589)
TITLE	Bolotin,A., Ehrlich,S.D. and Sorokin,A.
JOURNAL	Studies of genomes of dairy bacteria <i>Lactococcus lactis</i>
COMMENT	Sci. Aliments (2002) In press
	Contact: Sorokin A
	Genetique Microbienne
	INRA

```

tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
best homologue in strain IL1403 is deoD (99%)
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 1561.
Location/Qualifiers
1..1589
/organism="Lactococcus lactis subsp. cremoris"
/mol_type="genomic DNA"
/strain="MG1363"
/sub_species="cremoris"

```

[illegible]

RESULT 9	CF306721	169 bp	mRNA	linear	EST 15-AUG-2003
LOCUS	HDAL--04-K17.g1	CF306721	OshDACL1-overexpressing transgenic rice lambda phage		
DEFINITION	cDNA library 1 (HDAL1) <i>Oryza sativa</i> (japonica cultivar-group) cDNA clone HDAL--04-K17, mRNA sequence.				
ACCESSION	CF306721				
VERSION	CF306721.1	GI:33678482			
KEYWORDS	EST.				
SOURCE	<i>Oryza sativa</i> (japonica cultivar-group)				
ORGANISM	<i>Oryza sativa</i> (japonica cultivar-group)				
	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; <i>Oryza</i> .				
REFERENCE	1 (bases 1 to 169)				
AUTHORS	Kim, J. S., Jun, K. M., Cheong, P. J., Kim, M. J., Lee, T. H., Shin, Y. C., Song, S. I., Kim, J. K., Kim, Y. K. and Nahm, B. H.				

TITLE Large-scale Sequencing Analysis of Rice ESTs
JOURNAL Unpublished (2003)
COMMENT Contact: Nahm B.H.

Genomics and Genetics Institute, GreenGene Biotech Inc.; Division
of Bioscience and Bioinformatics, Myongji University
Yongin, Kyeonggi, Korea
Tel: 82 31 330 6193
Fax: 82 31 321 6355
Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

FEATURES

source

1. .169

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nackdong"

/db_xref="taxon:39947"

/clone="HDA1--04-K17"

/tissue_type="callus"

/dev_stage="proliferated callus on 2N6 media for 2 weeks"

/lab_host="E.coli SOLR"

/clone_lib="OSHDA1-overexpressing transgenic rice lambda

phage cDNA library I (HDA1)"

/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:

XhoI; Callus was treated with ABA(20um) for 1hour. cDNA

was inserted into lambda Uni-ZAP XR vector at 5' end with

EcoRI and 3' end with XhoI site. mRNA was derived from

rice Histone Deacetylase overexpression line."

ORIGIN

Query Match 23.5%; Score 169; DB 7; Length 169;

Best Local Similarity 100.0%; Pred. No. 3.4e-41;

Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 GGCGCAGTTCTGCGCGCAGCTAAACCTGCGCAGCTGCTTATCGGTATGGTGGTGCCTGCACC 336

Db 1 GGCGCAGTTCTGCGCGCAGCTAAACCTGCGCAGCTGCTTATCGGTATGGTGGTGCCTGCACC 60

Qy 337 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGGCGTATCGCTGACTTC 396

Db 61 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGGCGTATCGCTGACTTC 120

Qy 397 GACATGGTGGCTTAACCGCATGAGTGCAGCTAAGCATTGGGTATTGATG 445

Db 121 GACATGGTGGCTTAACCGCATGAGTGCAGCTAAGCATTGGGTATTGATG 169

RESULT 10

CF306860

DEFINITION HDAL1--05-A22.g1 OshDACL1-overexpressing transgenic rice lambda phage

cDNA library I (HDA1) Oryza sativa (japonica cultivar-group) cDNA

clone HDAL1--05-A22, mRNA sequence.

ACCESSION CF306860

VERSION CF306860.1 GI:33678621

KEYWORDS EST.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaeae; Oryza.

1 (Bases 1 to 169)

Kim, J.S., Jun, K.M., Cheong, P.J., Kim, M.J., Lee, T.H., Shin, Y.C.,

Song, S.I., Kim, J.K., Kim, Y.-K. and Nahm, B.H.

Large-scale Sequencing Analysis of Rice ESTs

Unpublished (2003)

Contact: Nahm B.H.

Genomics and Genetics Institute, GreenGene Biotech Inc.; Division

of Bioscience and Bioinformatics, Myongji University

Yongin, Kyeonggi, Korea

Tel: 82 31 330 6193

Fax: 82 31 321 6355

Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

Location/Qualifiers

1. .169

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nackdong"

/db_xref="taxon:39947"

/clone="HDA1--06-K08"

/tissue_type="callus"

/dev_stage="proliferated callus on 2N6 media for 2 weeks"

/lab_host="E.coli SOLR"

/clone_lib="OSHDA1-overexpressing transgenic rice lambda

phage cDNA library I (HDA1)"

/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nackdong"
/db_xref="taxon:39947"
/clone="HDA1--05-A22"
/tissue_type="callus"
/dev_stage="proliferated callus on 2N6 media for 2 weeks"
/lab_host="E.coli SOLR"
/clone_lib="OSHDA1-overexpressing transgenic rice lambda
phage cDNA library I (HDA1)"
/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; Callus was treated with ABA(20um) for 1hour. cDNA
was inserted into lambda Uni-ZAP XR vector at 5' end with
EcoRI and 3' end with XhoI site. mRNA was derived from
rice Histone Deacetylase overexpression line."

ORIGIN

Query Match 23.5%; Score 169; DB 7; Length 169;

Best Local Similarity 100.0%; Pred. No. 3.4e-41;

Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 277 GGCGCAGTTCTGCGCGCAGCTAAACCTGCGCAGCTGCTTATCGGTATGGTGGTGCCTGCACC 336

Db 1 GGCGCAGTTCTGCGCGCAGCTAAACCTGCGCAGCTGCTTATCGGTATGGTGGTGCCTGCACC 60

Qy 337 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGGCGTATCGCTGACTTC 396

Db 61 GATTCCAAAGTTAACCGCATCCGTTTTAAAGACCATGACTTTGGCGTATCGCTGACTTC 120

Qy 397 GACATGGTGGCTTAACCGCATGAGTGCAGCTAAGCATTGGGTATTGATG 445

Db 121 GACATGGTGGCTTAACCGCATGAGTGCAGCTAAGCATTGGGTATTGATG 169

RESULT 11

CF307418

DEFINITION HDAL1--06-K08.g1 OshDACL1-overexpressing transgenic rice lambda phage

cDNA library I (HDA1) Oryza sativa (japonica cultivar-group) cDNA

clone HDAL1--06-K08, mRNA sequence.

ACCESSION CF307418

VERSION CF307418.1 GI:33679179

KEYWORDS EST.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaeae; Oryza.

1 (Bases 1 to 161)

Kim, J.S., Jun, K.M., Cheong, P.J., Kim, M.J., Lee, T.H., Shin, Y.C.,

Song, S.I., Kim, J.K., Kim, Y.-K. and Nahm, B.H.

Large-scale Sequencing Analysis of Rice ESTs

Unpublished (2003)

Contact: Nahm B.H.

Genomics and Genetics Institute, GreenGene Biotech Inc.; Division

of Bioscience and Bioinformatics, Myongji University

Yongin, Kyeonggi, Korea

Tel: 82 31 330 6193

Fax: 82 31 321 6355

Email: bhnahm@gbio.com, bhnahm@bio.myongji.ac.kr.

Location/Qualifiers

1. .161

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nackdong"

/db_xref="taxon:39947"

/clone="HDA1--06-K08"

/tissue_type="callus"

/dev_stage="proliferated callus on 2N6 media for 2 weeks"

/lab_host="E.coli SOLR"

/clone_lib="OSHDA1-overexpressing transgenic rice lambda

phage cDNA library I (HDA1)"

/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:

XhoI; Callus was treated with ABA(20um) for 1hour. cDNA was inserted into lambda Uni-ZAP XR vector at 5' end with EcoRI and 3' end with XhoI site. mRNA was derived from rice Histone Deacetylase overexpression line."

ORIGIN

```

Query Match      20.8%; Score 150; DB 7; Length 161;
Best Local Similarity 99.4%; Pred. No. 2.8e-35;
Matches 161; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 284 TTTCGCCGACGTAACAACTCGCGACGTCGTTATCGGTATGGTGCCTGCACCGAATCCA 343
Db 1 TTTCGCCGACGTAACAACTCGCGACGTCGTTATCGGTATGGTGCCTGCACCGAATCCA 60

Qy 344 AAGTTAACCGCATCCGTTTAAAGACCATGATCTTTCGGCTATCGCTTCGACATGG 403
Db 61 AAGTTAACCGCATCCGTTTAAAGA-CATGACTTTTCGGCTATCGCTTCGACATGG 119

Qy 404 TGCCTAACGCACTAGATGCAGCTAAAGCACTGGGTATTGATG 445
Db 120 TGCCTAACGCACTAGATGCAGCTAAAGCACTGGGTATTGATG 161

```

RESULT 12

BH138974 921 bp DNA linear GSS 07-AUG-2001
 ENT0V94TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
 genomic, genomic survey sequence.

ACCESSION

BH138974

KEYWORDS

GSS.

SOURCE

Entamoeba histolytica

ORGANISM

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE

1 (bases 1 to 921)

AUTHORS

Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.

TITLE

Determination of clone end sequences from Entamoeba histolytica

JOURNAL

HM1:IMSS sheared DNA library (2001)

COMMENT

Unpublished (2001)
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjloftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library

Seq primer:

M13-Forward

Class:

shotgun

High quality sequence start:

2

High quality sequence stop:

563.

Location/Qualifiers

1. 921

source

/organism="Entamoeba histolytica"

/mol_type="genomic DNA"

/strain="HM1:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"
 /note="vector: pHOSt; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

ORIGIN

```

Query Match      16.9%; Score 121.8; DB 8; Length 921;
Best Local Similarity 53.2%; Pred. No. 2.9e-26;
Matches 304; Conservative 0; Mismatches 262; Indels 5; Gaps 2;

Qy 5 CTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTGTTCGATGCCAGCG 64
Db 128 CAACTCCACATATGCTGCTAAATACGGAGAGATTGCTGAACTGCTCTTATGGCTGGTG 187

Qy 65 ACCCGCTGCGTGGAGATATATTGCTGAAACTTTCCTTGAAGATGCCCGGGAAGTGAACA 124
Db 188 ATCCATTAAGAGTAAACCTTCTTGCAGACACTTATTGACTGATGTTGTTCAATAATA 247

Qy 125 ACCTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGGCCCAAAATTTTCGTAATGG 184
Db 248 GTGTTAGAGAGCAGTAGGATATACCTGATATTAAGAGGAGTGAAGTTATCAGTTCAAG 307

Qy 185 GTCACGCTGTGTGATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGATTCG 244
Db 308 CTCATGCTATGGGTATGCCATCTATTGGAATTTATGCATATGAATTTATTTCTATG 367

Qy 245 GCGTGAAGAAAATTATCCCGCTGGGTTCCTGTGGCGCAGTTCTGCCGACGTAACACTGC 304
Db 368 GAGTAAAGAGAAATTATTAGAAATTTGTTCTGCTGGAGCTTTTGATGAAGTTTAAACTTG 427

Qy 305 GCGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAAACCGCATCCGTTTA 364
Db 428 GAGATATGTTATGTAATGGAGCATGTTATGATTTCCAAATTTTGAAGACAATATGATA 487

Qy 365 AAGACCATGACTTTGCGCTATCGCTGCTGCTGACATTCGATGCGTAAACGCAAGTAGATCAG 424
Db 488 TTCACGTTAAATACTCATGTATTGCTGATTTCCAACTTTTCAGAGAAGAGTGTGATGCAG 547

Qy 425 CTAAGCACTGGTATGATGCTCGGTGGGTAACTGTTCTCGCTGACCTGTTCTACT 484
Db 548 CTGAAAACCTGGATATAGAT-ATAAAGTAAGGAATATATTTCGTGTAATTTTCTATG 606

Qy 485 CTCGGACGCGCAAAATGTTTCGAGCTGATGAAATAATACGGCAATTCGCGGTGGAATGG 544
Db 607 ATGATGGAGATCAC----TCCGGAGCATGGAAGAGTGGGAGTACTTGTGTGTAAGATG 662

Qy 545 AAGCGGCTGGTATCTACGGCGTCGCTGCAGA 575
Db 663 AAGCAGCTGCTCTTTATATGATTGCAGCAGA 693

```

RESULT 13

AZ680406/c

LOCUS

AZ680406

ENTGM40TF Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000

genomic, genomic survey sequence.

ACCESSION

AZ680406

VERSION

AZ680406.1

KEYWORDS

GSS.

SOURCE

Entamoeba histolytica

Entamoeba histolytica

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE

1 (bases 1 to 863)

AUTHORS

Loftus, B., Van Aken, S. and Fraser, C.

TITLE

Determination of clone end sequences from Entamoeba histolytica

JOURNAL

HM1:IMSS sheared DNA library

COMMENT

Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library

Seq primer: M13-Forward

Class: shotgun

```

FEATURES
source
    High quality sequence start: 26
    High quality sequence stop: 769.
    Location/Qualifiers
        1. .863

```

AUTHORS	JOURNAL
TITLE	COMMENT
<p>1. <i>Journal of the American Medical Association</i>, 1990; 263: 2761-2765.</p> <p>2. <i>Journal of the American Medical Association</i>, 1990; 263: 2766-2768.</p>	<p>1. <i>Journal of the American Medical Association</i>, 1990; 263: 2769-2771.</p> <p>2. <i>Journal of the American Medical Association</i>, 1990; 263: 2772-2774.</p>

Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
Determination of clone end sequences from *Entamoeba histolytica*
HMI:IMSS sheared DNA library (2001)
Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the *Entamoeba histolytica* HMI:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 804.

FEATURES

Location/Qualifiers

1. .861

/organism="Entamoeba histolytica"

/mol_type="genomic DNA"

/strain="HM1:IMSG"

/db_xref="taxon:5759"

/clone.lib="Entamoeba histolytica Sheared DNA"

/note=Vector: pHO51; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of *E. histolytica* C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. *Exp. Parasitol.* 77:450. The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v+ method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In *Genome Sequencing: A Practical Approach*, eds. M. Vaubin and B. Barrell. Oxford University Press, 1999).

ORIGIN

Query Match
14.1%; Score 101.8; DB 8; Length 861;

Qy	5	CTACCCACACATTAATGCAAGAAATGGCGGATTTCCGTGACGTAGTTTTGATGCCAGGCG	64
Db	466	CTACTCCACACAACACGCTAAATATAGCTGATTTTGTGATACAGTGTTAGTAGCTGGTG	407
Qy	65	ACCGCTGCGTGCAGATATATTGCTCAAACTTTCTTGAAGATGCCGTCGAAGTGACA	124
Db	406	ATCCACTAAGATGTAAATTTGTTGCACACACTTATTTGACTGATGTGTTCCAAGTCAATG	347
Qy	125	ACGTTCCGGGTGATCTGGGGCTTACC CGGTACTTACAAAGGCCGCAAAATTTTCCGTAATGG	184
Db	346	GAGTAAGAGGGGCATTAGGATNATCTGGATATACAAAGGAAAGGTTTCTATTCAAG	287
Qy	185	GTACCGGTGTGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGGATTCG	244
Db	286	CTCATGTAAGGGTATACCATCTATTGGAAATTTATGCATATGAATATTTAAATTTCTATG	227
Qy	245	CGGTGAAGAAAAATTATCCGGGTGGGTTCCTGTGGCGGAGTTCTCGCGCACGTAAACCTGC	304
Db	226	GAGTAAAGAGAAATTTATAGAAATTTGGTTCTGCTGGAGCTATTAGCTCTAAATTTAGATATTG	167
Qy	305	GCAGCGTGTATTCGGTATGGGTGCCCTGCACCGGATTCCAAAGTTAAACCGCATCCGTTTTA	364
Db	166	GAGATATTGTATAGCAACTGGTGGCTTTGTCAATAATAGTAATTTTATGCAACAATATAAC	107
Qy	365	AAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGGTCGGTAAACGCAGTAGATGCAG	424
Db	106	TCCCAGGAACTTCTGCCCAATAGCTGATTTATGAGTTAGTTTCATTAAGCTGTTGAGCTG	47
Qy	425	CTAAGCAGCTGGG	437

Query Match	15.3%	Score 110;	DB 8;	Length 863;
Best Local Similarity	51.9%;	Pred. No. 1.3e-22;		
Matches 248;	Conservative 0;	Mismatches 230;	Indels 0;	Gaps 0;
Qy	5	CTACCCACACATTAATGCGAGAAATGGCGCATTTCCGTGAGTGTATTTTGATGCCAGCG	64	
Db	547	CTACTCCACACAACACGCTAATATAGCTGATTTTGCTGTATACAGTGTATTAGTCTGGTG	488	
Qy	65	ACCGCTGCGTGCAGATATATTCTGAAACTTTTCTTGAAGATGCCGTGAAGTGAACA	124	
Db	487	ATCCACTAAGATGTAAATTTGTCGACACACTTATTTGCACTGATGTGTTTCAAGTCAATG	428	
Qy	125	ACGTTCCGGTATGCTCGGCTTCAACGGTACTTACAAAGGCCGCAAAATTTTCGTAATGG	184	
Db	427	GAGTAAGAGGGCATTAGGATATATCTGATATTACAAAGGAAAAAGGTTTCTATTCCAAG	368	
Qy	185	GTCAGGTGTGGTATCCGCTCCTGCTCCATCTACACAAAGAAGTCAATCCAGATTTCG	244	
Db	367	CTCATGTATGGGTATACCATCTATTGGAAATTTATGCATATGAATTAATTTAAATTTCTATG	308	
Qy	245	GCGTGAAGAAAAATTATCCGGTGGGTCTCTGTGGCGAGTTCTGCCGACGTAAACATGC	304	
Db	307	GAGTAAAGAGAAATTAATTAGAAATTTGGTCTGCTGGAGCTATTAGCTCTAAATTAGATATTG	248	
Qy	305	GCGACGTTCGTTATCCGTTATGGGTGCCTGCACCGATTCCAAAGTTAAACCGCATCCGTTTAA	364	
Db	247	GAGATATTGTCATAGCAACTGGTCTGTGCATATAGTAATTTTATGCACAATATAAAC	188	
Qy	365	AAGACCATGACTTTGCCGCTATCGCTGACTTCGATCTCGACATGGTGGCGTAACGCGATAGATCGAG	424	
Db	187	TCCCAGGAACAATTCGTGCCAATAGCTGATTATGATTTAGTTAGTTTCAATAAAGCTGTTCAAGCTG	128	
Qy	425	CTAAGCACTGGGTATTGATGCTCCGGTGGGTAACTGTTCTCCGCTGACCTGTTCTTA	482	
Db	127	TTGAAAAGATTGGAGCTTAAGTATGCAAGTAGGGAATGTTTTTCTAGTGATGTTTCTTA	70	

RESULT 14	BH166036	861 bp	DNA	linear	GSS 24-SEP-2001
LOCUS	ENTSV60TF	Entamoeba histolytica	Sheared	DNA	Entamoeba histolytica
DEFINITION	genomic, genomic survey sequence.				
ACCESSION	BH166036				
VERSION	BH166036.1	GI:15739474			
KEYWORDS	GSS.				
SOURCE	Entamoeba histolytica				
ORGANISM	Entamoeba histolytica				
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.				
	1 (bases 1 to 861)				

```
Db      46  TGAAGATGGAG 34

RESULT 15
AZ675191
LOCUS   ENT59TR Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000
DEFINITION genomic, genomic survey sequence.
ACCESSION   AZ675191
VERSION     AZ675191.1 GI:11812337
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
ORGANISM    Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1 (bases 1 to 864)
AUTHORS     Loftus, B., Van Aken, S. and Fraser, C.
TITLE       Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
JOURNAL     Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: b.loftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 21
            High quality sequence stop: 808.

FEATURES             source
     1..864
         /organism="Entamoeba histolytica"
         /mol_type="genomic DNA"
         /strain="HM1:IMSS"
         /db_xref="taxon:5759"
         /clone_lib="Entamoeba histolytica Sheared DNA"
         /note="Vector: pHOsi; Site 1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barell, Oxford University Press, 1999)."
```

```
QY      245  GCCTGAAGAAAAATTATCCG-CGTGGGTTCTGTGGCGCAGTTCTGCGCACGTAATAACTG 303
         |||||
Db      647  GAGTAAAGAGAAATTATAGAAATGGTTCTGCTGGAGCCCTTTGATGAAGTTTAAACTT 706
         |||||

QY      304  CGCGACGTCGTTATCGGTATGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGTTTT 363
         |||||
Db      707  GGAGATATTGTTATTGGAATGGGAGCATGTTATGATTTCCAAATTTGAAAGACAATATGAT 766
         |||||

QY      364  AAAGACCATGACTTTGCGCGCTATCGCTGACTTCGACATGCTGCTAACGCGAGTAGATGCA 423
         |||||
Db      767  ATTCCAGGTAATACTCATGTATTGCTGATTTCCAACTTTGCAGAGAAACAGTTGATGCA 826
         |||||

QY      424  GCTAAAAGCACTGG 436
         |||||
Db      827  GCTGAAAAAAGCTGG 839
         |||||

Search completed: July 22, 2005, 22:27:44
Job time : 4174 secs
```

```
QY      5  CTACCCACACATTAATGCAGAAATGGCGAATTCGTCGACGTAGTTTGTATGCCAGCG 64
         |||||
Db      407  CNACTCCACATATATGGTGCTAAATACGAGAGATTCGTAACCTGCTCTTATGGCTGGTG 466
         |||||

QY      65  ACCCGCTGCGGCGAAGTATATTCCTGAAACTTCCTTGAAGTCCCGTGAAGTGAACA 124
         |||||
Db      467  ATCCATTAAAGAGTAANAACCTTCTTCGCACACACTTATTTGACTGATGTTGTTCAATATAATA 526
         |||||

QY      125  ACGTTCGCGGTATCGTGGGCTTACCCGGTACTTACAAAGCCGCAAAATTTCCGTAATGG 184
         |||||
Db      527  GTGTTAGAGGACAGTAGGATATAGTATACAGGGAGTGAACTTATCAGTTCAAG 586
         |||||

QY      185  GTACGGTGTGGTATCCGTCCTCGCTCCATCTACACCAAGAACTGATCACCGATTTTCG 244
         |||||
Db      587  CTCATGTTATGGGTATGCCATCTATTGGAAATTTATGCATATGAATTATTAATTTCTATG 646
         |||||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:26:10 ; Search time 43 Seconds
(without alignments)
414.910 Million cell updates/sec

Title: US-10-035-300A-2
Perfect score: 1225
Sequence: 1 MATPHINAEAGDFADVLMPCDPLRAKYIAETFLFEDAREVNNVRGMLGFTGTGKGRKISV 239

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues
Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1162	94.9	243	4	US-09-489-039A-8802
2	1057	86.3	264	4	US-09-543-681A-6543
3	653	53.3	221	4	US-09-902-540-10882
4	607	49.6	244	4	US-09-107-532A-4839
5	591	48.2	237	4	US-09-134-000C-6013
6	576.5	47.1	236	4	US-09-583-110-2698
7	576.5	47.1	237	4	US-09-107-433-4570
8	572	46.7	243	3	US-09-134-001C-4239
9	538.5	44.0	227	4	US-09-634-238-214
10	245.5	20.0	298	4	US-09-489-039A-8031
11	227.5	18.6	265	4	US-09-543-681A-5709
12	139	11.3	236	4	US-09-107-532A-6816
13	136	11.1	238	4	US-09-489-039A-12428
14	131	10.7	242	4	US-09-543-681A-6066
15	130	10.6	232	3	US-09-347-878-18
16	130	10.6	232	4	US-09-711-164-397
17	122.5	10.0	236	4	US-09-134-000C-5104
18	116.5	9.5	308	4	US-09-328-352-4660
19	115	9.4	228	4	US-09-710-279-2936
20	115	9.4	228	4	US-09-710-279-3196
21	115	9.4	233	3	US-09-134-001C-3643
22	108	8.8	230	4	US-09-583-110-5212
23	108	8.8	239	4	US-09-107-433-3429
24	106	8.7	254	4	US-09-583-110-5019
25	103.5	8.4	330	4	US-09-949-016-11585
26	101	8.2	237	4	US-09-107-433-4653
27	91.5	7.5	435	4	US-09-248-796A-17067

28	89.5	7.3	390	4	US-09-452-937A-41	Sequence 41, Appl
29	89	7.3	218	4	US-09-134-000C-4009	Sequence 4009, Ap
30	89	7.3	639	4	US-09-252-991A-18903	Sequence 18903, A
31	87	7.1	830	4	US-09-252-991A-20619	Sequence 20619, A
32	85	6.9	607	4	US-09-540-236-2006	Sequence 2006, Ap
33	84.5	6.9	389	4	US-09-489-039A-14234	Sequence 14234, A
34	84.5	6.9	389	4	US-09-949-016-9488	Sequence 9488, Ap
35	84	6.9	546	4	US-09-396-154-34	Sequence 34, Appl
36	81.5	6.7	445	1	US-08-674-168-30	Sequence 30, Appl
37	81.5	6.7	445	3	US-08-985-908-20	Sequence 20, Appl
38	81.5	6.7	445	3	US-08-852-730-5	Sequence 5, Appli
39	81.5	6.7	445	3	US-08-985-916-12	Sequence 12, Appl
40	80.5	6.6	390	4	US-09-452-937A-20	Sequence 20, Appl
41	80	6.5	249	4	US-09-107-532A-6722	Sequence 6722, Ap
42	79.5	6.5	363	4	US-09-358-321C-12	Sequence 12, Appl
43	79.5	6.5	393	4	US-09-358-321C-6	Sequence 6, Appli
44	79.5	6.5	411	4	US-09-543-681A-7404	Sequence 7404, Ap
45	79.5	6.5	713	4	US-09-270-767-45101	Sequence 45101, A

ALIGNMENTS

RESULT 1
US-09-489-039A-8802
; Sequence 8802, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8802
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8802

Query Match	94.9%	Score 1162;	DB 4;	Length 243;
Best Local Similarity	94.6%	Pred. No. 8.6e-128;		
Matches 226;	Conservative 6;	Mismatches 7;	Indels 0;	Gaps 0;
QY	1	MATPHINAEAGDFADVLMPCDPLRAKYIAETFLFEDAREVNNVRGMLGFTGTGKGRKISV	60	
Db	5	MATPHINAEAGDFADVLMPCDPLRAKYIAETFLFEDAREVNNVRGMLGFTGTGKGRKISV	64	
QY	61	MHGVGIPSCSIYTKELITDFGVKKIIRVSGCGLVPHVKLRDVVVGACTDSKVNRI	120	
Db	65	MHGVGIPSCSIYTKELITDFGVKKIIRVSGCGLVPHVKLRDVVVGACTDSKVNRI	124	
QY	121	FKDHFADFDVNRNADAAKALGDARVGNLFSDLFYSPDGMFDMVMEKYGLGVE	180	
Db	125	FKDHFADFDVNRNADAAKALGDARVGNLFSDLFYSPDGMFDMVMEKYGLGVE	184	
QY	181	MEAGIYGVAAEPGAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGDKE	239	
Db	185	MEAGIYGVAAEPGAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGDKE	243	
RESULT 2				
US-09-543-681A-6543				
; Sequence 6543, Application US/09543681A				
; Patent No. 6605709				
; GENERAL INFORMATION:				
; APPLICANT: GARY BRETON				
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS				
; FILE REFERENCE: 2709.1002-001				

; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6543
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6543

Query Match 86.3%; Score 1057; DB 4; Length 264;
Best Local Similarity 87.2%; Pred. No. 2e-115;
Matches 205; Conservative 15; Mismatches 15; Indels 0; Gaps 0;

QY 1 MATPHINAENGDFADVLMFGDPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db MATPHINAENGDFADVLMFGDPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 86

QY 61 MGHGVGPSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIWGMGACTDSKNRIR 120
Db MGHGMGPSIYAKELITDFGVKKIIRVSGCGAVLPHVKLRDVIWGMGACTDSKNRLR 146

QY 121 FKQDFAAIAADFQWVRNAVDAKALGIDARVGNLFSADLFYSPDGENFDMVKYGIIGVE 180
Db FKQDFAAIAADFELVQNAVSAKAKDVKRVGNLFSADLFYSPDPEFDMVKYGIIGVE 206

QY 181 MEAAGYGVAAEFGAKALTCTVSDHIRTHTQTTAERQTTFNDMIKIALESVLL 235
Db MEAAGYGVAAEFGARALTCTVSDHIKGTQTTSEERQTTTFNEMIEIALESVLL 261

RESULT 3
US-09-502-540-10882
; Sequence 10882, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10882
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-502-540-10882

Query Match 53.3%; Score 653; DB 4; Length 221;
Best Local Similarity 58.2%; Pred. No. 3.6e-69;
Matches 124; Conservative 39; Mismatches 50; Indels 0; Gaps 0;

QY 19 MPGDPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISVNGHGVGPSIYTKELI 78
Db 1 MPGDPLRAKYISDRFLEGAREVTSVRNMLGFTGTFGRRRVSVNGHGMGPSIYATALI 60

QY 79 TDFGVKKIIRVSGCGAVLPHVKLRDVIWGMGACTDSKNRIRFKDHPFAIADPDVNRNA 138
Db 61 KTYGVRVIRIIRVSGCGALSTVDKVRREVIVATGAGTDSNNVRMLKGHDFAAIVADFTLARRA 120

QY 139 VDAKALGIDARVGNLFSADLFYSPDGENFDMVKYGIIGVEAAGYGVAAEFGAKAL 198
Db 121 MEAERENKVRAGPVFTSLDFYHPQQLNATLARNGVLAVENAVAGYGVAAESGARAL 180

QY 199 TICTVSDHIRTHTQTTAERQTTFNDMIKIALE 231
Db 181 GLLTVSDHIITGESLTPQERQTTFDEMIELALD 213

RESULT 4

US-09-107-532A-4839
; Sequence 4839, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4839:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...244
; SEQUENCE DESCRIPTION: SEQ ID NO: 4839:
US-09-107-532A-4839

Query Match 49.6%; Score 607; DB 4; Length 244;
Best Local Similarity 52.8%; Pred. No. 1e-62;
Matches 122; Conservative 40; Mismatches 67; Indels 2; Gaps 2;

QY 5 HINAENGDFADVLMFGDPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISVNGHG 64
Db 13 HIEAKEGEIADKLLFGDPLRAKYIAETFLDPCVNCYNQVRGMLGFTGKYKGERISVQGTG 72

QY 65 VGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIWGMGACTDSKNRIRFKDH 124
Db 73 MGNPSATIYAHLEIQSYGVKKLRVGTGALSKDVHVRDLVLAQGAATSSSMIEKQFQAF 132

QY 125 DFAAIAADFQWVRNAVDAKALGIDARVGNLFSADLFYSPD-GEMFDMVKYGIIGVEMEA 183
Db 133 HPPPIISDFNLLKAYEIAKEKGYTVHGVNLSVSDSFYKDDLTETFQALAE-LGVLGVEMEA 191

QY 184 AGYGVAAEFGAKALICTVSDHIRTHTQTTAERQTTFNDMIKIALESVL 234
Db 192 AALYILGAKYHVQTLMTVSDHLITGEETTAABROSTFNMIEVGLTETAI 242

RESULT 5
US-09-134-000C-6013
; Sequence 6013, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6013
; LENGTH: 237
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6013
Query Match 48.2%; Score 591; DB 4; Length 237;
Best Local Similarity 50.9%; Pred. No. 7,5e-61;
Matches 117; Conservative 42; Mismatches 71; Indels 0; Gaps 0;
Qy 5 HINAMGDFADVLMMPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYTKGRKISVMGHG 64
Db 6 HIEAKPGIADKILLPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYTKGRKISVMGHG 65
Qy 65 VGIPSCSYITKELITDFGVKKIIRVSGAVLPHVKLRDGVVIGMGACTDSKVNIRFKDH 124
Db 66 MGMPISAGIYAHELINSYOVKKIIRVGTGCSISEKVNVRRELVLQAATAATNSNIRNDPQY 125
Qy 125 DPAIADPDMVRNADAKALIGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVEMEA 184
Db 126 DFPQIASFDLLKAYHAKELGTMTHVGNVLSDDVFKSDLDVPRIGRYGVLAEMEA 185
Qy 185 GIYGVAAEFAGKALTICTVSDH-IRTHEQTAAERQTTFNDMIKIALESVL 234
Db 186 MLYLAQFDVEALAMTVSDSLVTGEETAAERQTTFNEMIEVGLTAI 235
RESULT 6
US-09-583-110-2698
; Sequence 2698, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 2698
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-2698
Query Match 47.1%; Score 576.5; DB 4; Length 236;
Best Local Similarity 49.8%; Pred. No. 3.7e-59;
Matches 115; Conservative 44; Mismatches 71; Indels 1; Gaps 1;
Qy 5 HINAMGDFADVLMMPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYTKGRKISVMGHG 64

Db 4 HIAAQGEIADKILLPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYTKGRKISVMGHG 63
Qy 65 VGIPSCSYITKELITDFGVKKIIRVSGAVLPHVKLRDGVVIGMGACTDSKVNIRFKDH 124
Db 64 MGMPISAGIYARELIVDYGKKIIRVGTAGSLNEEVHRELVLQAATAATNSNIRNDPQY 123
Qy 125 DPAIADPDMVRNADAKALIGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVEMEA 184
Db 124 DFPQIASFDLLKAYHAKELGTMTHVGNVLSDDVFKSDLDVPRIGRYGVLAEMEA 183
Qy 185 GIYGVAAEFAGKALTICTVSDH-IRTHEQTAAERQTTFNDMIKIALESVL 234
Db 184 ALYLAQHYVDALAMTISDSLVNPDDETTAERQNTFTDMKVGLETLI 234
RESULT 7
US-09-107-433-4570
; Sequence 4570, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4570:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...237
; SEQUENCE DESCRIPTION: SEQ ID NO: 4570:
US-09-107-433-4570
Query Match 47.1%; Score 576.5; DB 4; Length 237;
Best Local Similarity 49.8%; Pred. No. 3.8e-59;
Matches 115; Conservative 44; Mismatches 71; Indels 1; Gaps 1;
Qy 5 HINAMGDFADVLMMPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYTKGRKISVMGHG 64

```
Db 5 HIAAQGEIADKILLPGDPLRAKFAENFLGDVACFNEVRNMFYGTGTYGKHRVSMGTG 64
Qy 65 VGPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVIGMGACTDSKVNRIKPDH 124
Db 65 MGMPISIIYARELIVDGVKKLIRVAGAGSLNEEVRHVLVLAQAAATNSIVRNDMPQY 124
Qy 125 DFAAIAIDFDMVRNADAAKALGDARVGNLFSADLFYSPDGMFDMVMEKYGILGVEEAA 184
Db 125 DFQIASFDLLDKAYHIAKELGTHMTVGNVLSDFVSNFYFKNIELGKKGAVENEAA 184
Qy 185 GIYGVAAEFAGAKALTICTVSDH-IRTHEQTAAERQTTFNMDMKIALESVL 234
Db 185 ALYYLAAQYHVDALAIMTISDSLVNPDETTABERQNTFTDMMKVGLTILI 235

RESULT 8
US-09-134-001C-4239
; Sequence 4239, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC-007
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-11-08
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4239
; LENGTH: 243
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4239

Query Match 46.7%; Score 572; DB 3; Length 243;
Best Local Similarity 53.2%; Pred. No. 1.3e-58;
Matches 123; Conservative 30; Mismatches 76; Indels 2; Gaps 2;

Qy 2 ATPHINAEAGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISVM 61
Db 11 STPHINPVGKIAKTVLMPGDPPLRAQYIAENFLENVEQFTVRNMFYGTGTYGKEVSVM 70
Qy 62 GHGVGPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVIGMGACTDSK-VNRI 120
Db 71 GSGMGIPSGIYSELYHFDFDVTIIRVSCGALQEDVNLVDYIIAQAASTNSNYVDQFN 130
Qy 121 FKDHDFAAIADFDMVRNADAAKALGDARVGNLFSADLFYSPDGMFDMVMEKYGILGVE 180
Db 131 IPGH-FAPIADFNLVAKAKAABEIGAISHVGNVLSDDTFYNADSTFNDSWKWGIIGIE 189
Qy 181 MEAAGIYGVAAEFAGAKALTICTVSDHIRTHTHEQTAAERQTTFNMDMKIALE 231
Db 190 MESAGLYLNAIHANKALGIFTVSDHILRDEATSAERQTSFTQMMBEIALE 240

RESULT 9
US-09-634-238-214
; Sequence 214, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.
; APPLICANT: Lubbers, Mark W.
; APPLICANT: Dekker, James
; APPLICANT: Christenson, Anna C.
; APPLICANT: Holland, Rose
; APPLICANT: O'Toole, Paul W.
; APPLICANT: Reid, Julian R.
; APPLICANT: Coolbear, Timothy
```

```
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them and methods for using them.
; FILE REFERENCE: 11000.1043U1
; CURRENT APPLICATION NUMBER: US/09/634,238
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 214
; LENGTH: 227
; TYPE: PR1
; ORGANISM: Lactobacillus rhamnosus
US-09-634-238-214

Query Match 44.0%; Score 538.5; DB 4; Length 227;
Best Local Similarity 51.1%; Pred. No. 1e-54;
Matches 116; Conservative 39; Mismatches 69; Indels 3; Gaps 3;

Qy 1 MATPHINAEAGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db 1 MST-HIDAPKGAIAADVLLPGDPLRAQYIAEHLEKAVRYNTVRNAPGYTGTTEGRRISV 59
Qy 61 MGHGVGPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVIGMGACTDSKVNRI 120
Db 60 QATGMGIPSGIYVNELIQDYGVKTLIRVGTAGMGSDVKVRDVLVQGSSTDSIVLNT 119
Qy 121 F-KDHDFAAIAADFDMVRNADAAKALGDARVGNLFSADLFYSPDGMFDMVMEKYGILV 179
Db 120 FGAGMYFAPIAIDFQILREAAANLADAGALRYHGVNVLGEDRFYNDMDRQKLID-YGV 178
Qy 180 EMEAGIYGVAAEFAGAKALTICTVSDHIRTHTHEQTAAERQTTFNMDI 226
Db 179 EMETPALYLLAAKFAQAALSILTVSNHLITGETTAAERQTSFNMDI 225

RESULT 10
US-09-489-039A-8031
; Sequence 8031, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 2709.2004001
; CURRENT FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8031
; LENGTH: 298
; TYPE: PR1
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8031

Query Match 20.0%; Score 245.5; DB 4; Length 298;
Best Local Similarity 28.8%; Pred. No. 3.5e-20;
Matches 72; Conservative 45; Mismatches 102; Indels 31; Gaps 8;

Qy 5 HINAEAGDP--ADVVLMPGDPPLRAKYIAETF-----LEDAREVNNVRGMLGFTGTYGKRK 57
Db 53 LGLTKNDLQAGATLAIIVPGDPERVEKIAALMDKPKLASHREFTSWRAEL-----DGKP 106
Qy 58 ISVMGHGVGPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVIGMGACTDSKVN 117
Db 107 VIVCGTIGGFGSTIAVEEL-AQLGVRTFLRIGTTGAIQHIHNVGDVLVTTGVSRLDGS 165
Qy 118 RIRFKDHDFAAIAADFDMVRNADAAKALGDARVGNLFSADLFYSPDGMFV----- 170
Db 166 -LHFAPMEFPADVADFACTTALVEAASIGATTHIGVTASSDTFY-PQERYDTFSGRWS 223
Qy 171 -----MEKY---GILGVEAAGIYGVAAEFAGAKALTICTVSDHIRTHTHEQTAAERQTT 222
Db 224 RFGKSMEEQAMGVNMEYEMESATLLTMCASQGLRAGMAGVIVNRTQQEIPNAETMKQTE 283
```



```
Qy 223 NDMIKIALES 232
; :|||:
Db 284 SHAVKIVVEA 293

RESULT 11
US-09-543-681A-5709
; Sequence 5709, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5709
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5709

Query Match 18.6%; Score 227.5; DB 4; Length 265;
Best Local Similarity 26.4%; Pred. No. 3.8e-18;
Matches 65; Conservative 54; Mismatches 104; Indels 23; Gaps 8;

Qy 5 HINAEKMGDF--ADVVLMPGDLRAKIAETPLEDAENVNVRGMLGFTGTYGKRKISVNG 62
; :|||:
Db 19 HGLTKNDLQATLAIIVPCDPKRVKEIAK-LMDNPVHLASLREYTSWRGEIDGKAVIVCS 77
; :|||:
Qy 63 HGVGTPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIWICMGACTDSKNRIRPK 122
; :|||:
Db 78 TGIGGPSTSIKVEEL-AQLGRTFURIGTGTGAIQHIINVGDLIVTAAVRLDGAS-LHPA 135
; :|||:
Qy 123 DHDFAAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGENFDV----- 170
; :|||:
Db 136 PWEFPVSDFCFNMALYKAAKONGSTVHVGVTTASSDTFY-PQERYDYTYTGRVVRFRKGS 194
; :|||:
Qy 171 ---MEKYGILGVEMEAAGIYGVAAEFGAKALTICTVSDHIRTQTTAAE-RQTTFNMI 226
; :|||:
Db 195 MKWQEMGMVMEYSATLLTMCASQGLRAGMAGVIYN-RTQOSIPDAELLKKTENNAL 253
; :|||:
Qy 227 KIALES 232
; :|||:
Db 254 GIVIEA 259

RESULT 12
US-09-107-532A-6816
; Sequence 6816, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:

Qy 223 NDMIKIALES 232
; :|||:
Db 284 SHAVKIVVEA 293

APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6816:
SEQUENCE CHARACTERISTICS:
LENGTH: 236 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...236
SEQUENCE DESCRIPTION: SEQ ID NO: 6816:
US-09-107-532A-6816

Query Match 11.3%; Score 139; DB 4; Length 236;
Best Local Similarity 23.8%; Pred. No. 7.5e-08;
Matches 48; Conservative 37; Mismatches 83; Indels 34; Gaps 5;

Qy 45 GMLGFTGYTKRKISVMGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDV 104
; :|||:
Db 36 GALFISGLNHEVIVVRSGIGKVLASITTSLLQQYGVNVINTSGAGGIGEGIQVGDI 95
; :|||:
Qy 105 VIGMGACTDSKNRIRFKDHFDAI-----ADFDVNRNAVDAAKALGIDA 149
; :|||:
Db 96 VIS-----DKVAVFDADATGFGVNPGLPGMPLYEASTYLRSEMVAKAATNLNA 146
; :|||:
Qy 150 RVGNLFSADLFYSPDGENFDVMEKY-GILGVEMEAAGIYGVAAEFGAKALTICTVSDHIR 208
; :|||:
Db 147 KEGLVITGDTFVDSPKIKELITNPPEALACEMAGVAGTARQFNIPFLIVRAMSD--- 203
; :|||:
Qy 209 THEQTAAERQT-TFNDMIKIA 229
; :|||:
Db 204 -----TADHSATQSFDFIEDA 220

RESULT 13
US-09-489-039A-12428
; Sequence 12428, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12428
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12428

Query Match 11.1%; Score 136; DB 4; Length 238;
Best Local Similarity 23.0%; Pred. No. 1.7e-07;
Matches 53; Conservative 39; Mismatches 102; Indels 36; Gaps 6;
```

```
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-347-878-18

Query Match      10.6%; Score 130; DB 3; Length 232;
Best Local Similarity 21.3%; Pred. No. 8.3e-07;
Matches 49; Conservative 43; Mismatches 102; Indels 36; Gaps 6;

Qy 32 TFLEDA---REVNNVRGMLGFTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIR 88
Db 20 TLLRDKTIENRQTIYIGSEIYTGQHLGVDVALLKSGIGKVAAAGATLLERQCPDVIIN 79
Qy 89 VGSCGAVLPHVKLRDVIIVGMGACTDSKVNIRIRFKDHFDAAI-----ADFD 133
Db 80 TGSAGGLASTLKVGDIVVS-----DEARYHDADVTAFGYEVGQLPGCPAGFKADEK 130
Qy 134 MVRNADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKY-GILGVMEAAAGIYGVAAE 192
Db 131 LVAAAECAIKALDNLAVRGLIVSGDAFINGSVGLAKIRHNPQQAIAVEMEATAIAHVCHN 190
Qy 193 FGAKALITCTVSD-----HIRTHEQTAAERQTTFNDMIKIALESVLLG 236
Db 191 FKVPFVVVRAISDVADQQSHLSFEFLAVARQSTL--MVENLVQNLRG 238

RESULT 14
US-09-543-681A-6066
; Sequence 6066, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543.681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6066
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6066

Query Match      10.7%; Score 131; DB 4; Length 242;
Best Local Similarity 22.8%; Pred. No. 6.8e-07;
Matches 47; Conservative 38; Mismatches 85; Indels 36; Gaps 5;

Qy 49 FTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIRVGSCGAVLPHVKLRDVIIVGM 108
Db 41 YTGKINGVDVALLKSGIGKVAAAGTTLLEHFRPDVIVNTGVSAGGLDAKLVNGDIVVS- 99
Qy 109 GACTDSKVNIRIRFKDHFDAI-----IADFMMVRNADAAKALGIDARVGN 153
Db 100 -----TEVRVYHDADVTAFGYEPFGQWACPPAFIADPKLVNTAQECIGSLKLVNVRGL 151
Qy 154 LFSADLFYSPDGMFDMVEKY-GILGVMEAAAGIYGVAAEFGAKALITCTVSD-----H 206
Db 152 ICSGDAFINGAEPLARIRRTPEVVAVEMESTAGHCHQDPDTPFVVVRAISDVADKESH 211
Qy 207 IRTHEQTAAERQ-----TFENDMIK 227
Db 212 LSFDFELSVAAQQSSSLVTTMLDKLK 237

RESULT 15
US-09-347-878-18
; Sequence 18, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
; FILE REFERENCE: 25885-1651
; CURRENT APPLICATION NUMBER: US/09/347.878C
; CURRENT FILING DATE: 1999-07-06
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 18
; LENGTH: 232
```

```
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-347-878-18

Query Match      10.6%; Score 130; DB 3; Length 232;
Best Local Similarity 21.3%; Pred. No. 8.3e-07;
Matches 49; Conservative 43; Mismatches 102; Indels 36; Gaps 6;

Qy 32 TFLEDA---REVNNVRGMLGFTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIR 88
Db 14 TLLRDKTIENRQTIYIGSEIYTGQHLGVDVALLKSGIGKVAAAGATLLLEHCKPDVIIN 73
Qy 89 VGSCGAVLPHVKLRDVIIVGMGACTDSKVNIRIRFKDHFDAAI-----ADFD 133
Db 74 TGSAGGLAPTILKVGDIVVS-----DEARYHDADVTAFGYEVGQLPGCPAGFKADDK 124
Qy 134 MVRNADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKY-GILGVMEAAAGIYGVAAE 192
Db 125 LIAAAEACIAELNLNAVRLIVSGDAFINGSVGLAKIRHNPQQAIAVEMEATAIAHVCHN 184
Qy 193 FGAKALITCTVSD-----HIRTHEQTAAERQTTFNDMIKIALESVLLG 236
Db 185 FNVFVVVRAISDVADQQSHLSFDEFLAVAAKQSSL--MVESLVQKLAHG 232

Search completed: July 18, 2005, 21:41:28
Job time : 44 secs
```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:40:01 ; Search time 155 Seconds
(without alignments)
598.561 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225

Sequence: 1 MATPHINAEAGDFADVLMPL.....TTFNDMIKIALESVLLGDKE 239

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 398188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCT05_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US10F_PUBCOMB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1222	99.8	239	17 US-10-891-096-4	Sequence 4, Appli
2	693.5	56.6	233	16 US-10-772-272A-16	Sequence 16, Appl
3	675	55.1	233	9 US-09-895-913A-148	Sequence 148, App
4	663	54.1	231	15 US-10-335-977-8188	Sequence 8188, Ap
5	593	48.4	229	15 US-10-335-977-8187	Sequence 8187, Ap
6	576.5	47.1	237	18 US-10-617-320-4570	Sequence 4570, Ap
7	575.5	47.0	236	17 US-10-472-928-1598	Sequence 1598, Ap
8	572	46.7	243	18 US-10-724-972A-3878	Sequence 3878, Ap
9	538.5	44.0	227	15 US-10-264-213-122	Sequence 122, App
10	485.5	37.3	203	15 US-10-424-599-202583	Sequence 202583,
11	160.5	13.1	262	14 US-10-156-761-14181	Sequence 14181, A

ALIGNMENTS

RESULT 1
; Sequence 4, Application US/10891096
; Publication No. US20050074857A1
; GENERAL INFORMATION:
; APPLICANT: Araki, Tadashi
; APPLICANT: Miyake, Hitoki
; APPLICANT: Oikawa, Tooshihiro
; TITLE OF INVENTION: Method for Producing a Pyrimidine Nucleoside Compound
; FILE REFERENCE: 018765-161
; CURRENT APPLICATION NUMBER: US/10/891,096
; PRIOR FILING DATE: 2004-07-15
; PRIOR APPLICATION NUMBER: JP 2003-199175
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-891-096-4

Query Match	99.8%	Score 1222;	DB 17;	Length 239;
Best Local Similarity	99.6%	Pred. No. 8.9e-127;		
Matches 238;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MATPHINAEAGDFADVLMPLGDPPLRAKYIAETFLDAREVNNVRGMLGFTTYGKRKTSV	60	
Db	1	MATPHINAEAGDFADVLMPLGDPPLRAKYIAETFLDAREVNNVRGMLGFTTYGKRKTSV	60	
Qy	61	MHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIHNGACTDSKVNRI	120	
Db	61	MHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIHNGACTDSKVNRI	120	

QY 121 FKQDFAAIAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVE 180
Db 121 FKQDFAAIAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVE 180
QY 181 MEAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDK 239
Db 181 MEAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDK 239

RESULT 2

US-10-772-272A-16
; Sequence 16, Application US/10772272A
; Publication No. US20040166575A1
; GENERAL INFORMATION:
; APPLICANT: Ajinomoto Co., Inc.
; TITLE OF INVENTION: Inosine producing bacterium belonging to the genus
; TITLE OF INVENTION: Bacillus and method for producing inosine
; FILE REFERENCE: US-108
; CURRENT APPLICATION NUMBER: US/10/772,272A
; CURRENT FILING DATE: 2002-02-06
; PRIOR APPLICATION NUMBER: JP 2003-37760
; PRIOR FILING DATE: 2003-02-17
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-772-272A-16

Query Match 56.6%; Score 693.5; DB 16; Length 233;
Best Local Similarity 58.1%; Pred. No. 3.6e-68;
Matches 133; Conservative 41; Mismatches 54; Indels 1; Gaps 1;

QY 5 HINAEMGDFADVLMPCGDLRAKYAETFLDAREVNNVRGMLGFTGTYKGRKISVNGHG 64
Db 4 HIGAEGQIADTVLLPGDPLRAKYAETYLENVCEYNEVRGMYGFTGTYKGRKISVQGTG 63
QY 65 VGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVVIGMGACTDSKVNRIKDH 124
Db 64 MGVPISIIYVNELIQSDVQNLIRVSGGAIRKDKVRDVLAMTSDTSQMNRAFGSV 123
QY 125 DFAAIAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVEEAA 184
Db 124 DFACADFELLKNAYDAKDKGVPTVGSVFTADQFTNDSQI-EKLAQVGLGVEMETT 182
QY 185 GIYGVAAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNMDIKIALESV 233
Db 183 ALYTLAAKHGRKALSILTVSDHVLTBETTAERQTTTFHDMIEVALHSV 231

RESULT 3

US-09-895-913A-148
; Sequence 148, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148

; LENGTH: 233
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-148
Query Match 55.1%; Score 675; DB 9; Length 233;
Best Local Similarity 55.0%; Pred. No. 4.1e-66;
Matches 126; Conservative 42; Mismatches 61; Indels 0; Gaps 0;
QY 3 TPHINAEMGDFADVLMPCGDLRAKYAETFLDAREVNNVRGMLGFTGTYKGRKISVNG 62
Db 2 TPHINAKIGDFYPCQLCGDPLRVSIYAKFLQDAKEITNVRNMLGFSYKGRKISVNG 61
QY 63 HGVIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVVIGMGACTDSKVNRIKDH 122
Db 62 HMGIACTIYVTELLIKTVQKELLRAIGTCGALSPKVLKDIIMATGASTDSKTNVRFL 121
QY 123 DHFPAIAIDFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGLGVE 182
Db 122 NHDLSATPDPELSIRAYQTAIRGLGIDLVKGVNFSDFPFYSFETHAFDLMAKYNHLAEME 181
QY 183 AAGIYGVAEFGAKALTICTVSDHIRTHTQTTAAERQTTFNMDIKIALE 231
Db 182 AAGLYATAMELNAKALCLCSVDHLITKEALSPKRVESPDNMIILALE 230

RESULT 4

US-10-335-977-8188
; Sequence 8188, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 8188:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 231 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...231

SEQUENCE DESCRIPTION: SEQ ID NO: 8188;
US-10-335-977-8188

Query Match 54.1%; Score 663; DB 15; Length 231;
Best Local Similarity 53.7%; Pred. No. 8,6e-65;
Matches 123; Conservative 44; Mismatches 62; Indels 0; Gaps 0;
Qy 3 TPHNAEAGDPAVVLMPGDPRLAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISVMG 62
Db 2 TPHNAKIGDYPQCLCGDPLRVSYIAKNFLQDAKEITNVRNMLGFGSKYKGKISLMG 61
Qy 63 HGVIPSCSIYTKELITDFGVKKIIRVSCCAVLPVHVKLRDVIOMGACTDSKVNRFK 122
Db 62 HGMGIASCTIYVTELKTYQVKKELRIGTCGAISPKVGLKDIWMATGASTDSKTRVRFL 121
Qy 123 DHDFAAADDFMVNRNAVDAKALGIDARVGNLFSADLFPSPDGEFVMEKYGILGVEME 182
Db 122 NHDLSATPDFLSLRAYQTAKRLGIDLKIGNVFSDFYSFETHAFGLMAQYNHLAIME 181
Qy 183 AAGYGVAAEFGAKALITCTVSDHRTHEQTAAERQTFNDMIKIALE 231
Db 182 AAGLYATAMELNKALCLCSVDHLITKEALSPKRIESFDNMITLALE 230

RESULT 5

US-10-335-977-8187
; Sequence 8187, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
RELATING TO HELICOBACTER PYLORI FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: Windows NT 4.0
SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977
FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002
FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: GTN-018
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 8187:

SEQUENCE CHARACTERISTICS:

LENGTH: 229 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...229

SEQUENCE DESCRIPTION: SEQ ID NO: 8187:

US-10-335-977-8187

Query Match 48.4%; Score 593; DB 15; Length 229;
Best Local Similarity 52.4%; Pred. No. 5e-57;
Matches 110; Conservative 40; Mismatches 60; Indels 0; Gaps 0;
Qy 22 DPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISVMGCVGIPSCSIYTKELITDF 81
Db 20 DPFVSYIARYLLQDAKEITNVRNMLGFGSKYKGKISLMGHGMIASCTIYVTELKTY 79
Qy 82 GVKKIIRVSGCAVLPVHVKLRDVIIVGACATDSKVNRIKPDHFAAIAIDFDMVRNAVDA 141
Db 80 QVKELLRIGTCGALSPKVGLKDIWMATGASTDSKTRVRFLNHLDSATPDFLSLRAYQT 139
Qy 142 AKALGIDARVGNLFSADLFPSPDGEFVMEKYGILGVEMEAAAGYGVAAEFGAKALITC 201
Db 140 AKRLGIDLKIGNVFSDFYSFETHAFGLMAQYNHLAIMEAAGLYATAMELNKALCLC 199
Qy 202 TVSDHRTHEQTAAERQTFNDMIKIALE 231
Db 200 VSDHLITKEALSPKRIESFDNMITLALE 229

RESULT 6

US-10-617-320-4570

; Sequence 4570, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID

SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE

THERAPEUTICS

NUMBER OF SEQUENCES: 5206

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: <Unknown>

OPERATING SYSTEM: <Unknown>

SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/617,320

FILING DATE: 10-Jul-2003

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/107,433

FILING DATE: 30-Jun-1998

APPLICATION NUMBER: 60/ 085131

FILING DATE: May 12, 1998

APPLICATION NUMBER: 60/051553

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 4570:

SEQUENCE CHARACTERISTICS:

LENGTH: 237 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...237

Db 1 MST-HIDAPKGAIDVVLPGDPLRAQVIAEHFLEKAVRYNVRNFAFGYTGTFEGRRISV 59
Qy 61 MGHGVIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHKLRDVVVGMACTDSKVNIR 120
Db 60 QATGWGIPSIYVNELIODYGVKTLIRVGTAGGMSDKVRDVLVQSSSTDSSVLNT 119
Qy 121 F-KHDFAAIADFDVNRNADAAKALGIDARVGNLFSADLFSYSPDGMFDMVEKYGILGV 179
Db 120 FGAGMYFAPADFDQLLRANLADAGALRYHVGNVLDGDFYNDEMQRKLD-YGVILAT 178
Qy 180 EMEAGIYGVAAEFCAKALITCTVSDHIRTHTQTTAAERQTTFNMDI 226
Db 179 EMETPALYLLAAKFAQAALSILTVSNHLITGEETTAQERQTSFNMDI 225

RESULT 10

US-10-424-599-202583
; Sequence 202583, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 202583
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_24958C.1.pep
US-10-424-599-202583

Query Match 37.3%; Score 456.5; DB 15; Length 203;
Best Local Similarity 45.9%; Pred. No. 5.8e-42;
Matches 106; Conservative 24; Mismatches 66; Indels 35; Gaps 3;
Qy 2 ATPHINAEIMGDFADVLMFGDPLRAKYIAETFELEDAEVRNVRGMLGFTGYTKGRKISVM 61
Db 4 STPHINPVGVIKATVLMFGDPLRAQYIAENFLENVEQFNTVRNMFYGTGYTKGEVSM 63
Qy 62 GHGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHKLRDVVVGMACTDSK-VNRIR 120
Db 64 GSGMGIPSGIYSELVYHFFDVTIIRVSGCALQEDVNLVDVIAQAASNTSNVVDQFN 123
Qy 121 FKDHDFAAIADFDVNRNADAAKALGIDARVGNLFSADLFSYSPDGMFDMVEKYGILGVE 180
Db 124 IPGH-FAPADFNLVAKAKKADEIGALSHV----- 153
Qy 181 MEAGIYGVAAEFCAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALE 231
Db 154 ---GLVNLNATHANKALGIFTVSDHILRDEATSAERQTSFTQWMEIALE 200

RESULT 11

US-10-156-761-14181
; Sequence 14181, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA SATOSHI
; APPLICANT: IKEDA HARUO
; APPLICANT: ISHIKAWA JUN
; APPLICANT: HORIKAWA HIROSHI
; APPLICANT: SHIBA TADAYOSHI
; APPLICANT: SAKAKI YOSHIYUKI
; APPLICANT: HATTORI MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14181
; LENGTH: 262
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14181

Query Match 13.1%; Score 160.5; DB 14; Length 262;
Best Local Similarity 30.3%; Pred. No. 5.7e-09;
Matches 56; Conservative 31; Mismatches 91; Indels 7; Gaps 5;

Qy 13 PADVVLMPGDPLRAKYIAETFELEDAEVRNVRGMLGFTGYTKGRKISVMGHGVIPSCSI 72
Db 19 YAVVV---GDPAARAAVA-ALLDGAEEVSHREYRVFRGAWKGLPVTVASHGVGPGAIL 74
Qy 73 YTKELITDFGVKKIIRVSGCAGVLPVHKLRDVVVGMACTDSKVNIRIRFKDHDFAAIADF 132
Db 75 LFQEL-ADAGVTFVFRGTAGAKPGIGDGLVIAEAAVRDDGVTO-QLLPAAYPAVSAP 132
Qy 133 DMVRNADVAKALGIDARVGNLFSADLFSYSPDGMFDMVEKYGILGVMEAGIYGVAAE 192
Db 133 EAVLALQRAARETGAPHHRGIVVWRAAP-QDGLIPLDGYARAGLAAIEMELSALLVVTASL 191
Qy 193 FGAKA 197
Db 192 RGLVA 196

RESULT 12

US-10-282-122A-77737
; Sequence 77737, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77737
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-77737

Query Match      11.9%; Score 146; DB 15; Length 233;
Best Local Similarity 23.8%; Pred. No. 2e-07;
Matches 54; Conservative 39; Mismatches 98; Indels 36; Gaps 6;

Qy 32 TFLDAREVNVRGMLG---FTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIR 88
Db 14 TLLRDRIENRQTLARAGCEIYTGQLNGIDVALLKSGIKVAAAMGTTLLLEHCQPDIVN 73
Qy 89 VSGGAVLPHVKLRDVGVMGACTDSKVNRIKPDHFAA-----IADPD 133
Db 74 TGSAGGLDSSLKVGDIVVS-----NEVRHDADVTAFGEYPGQMGCPAFAVADE 124
Qy 134 MVRNAVDAAKALGIDARVGNLFSADLFYSPDGEFDMVEKY-GILGVMEAAAGIYGVAAE 192
Db 125 LIALAENCIQQLKLNARVLGICSDAFINGAEPLARIRAAFPPTVAAMEAAIGHVCYL 184
Qy 193 FGAKALTICTVSD-----HIRTQTTAAERQTTFNDMIKIALESV 233
Db 185 FNTPFVVVRAISVDVADQASHLSFEFLVVAAKQSTL--MIKAMLTTL 229

RESULT 13
US-10-282-122A-57936
; Sequence 57936, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57936
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Enterococcus faecium
US-10-282-122A-57936

Query Match      11.3%; Score 139; DB 15; Length 230;
Best Local Similarity 23.8%; Pred. No. 1.1e-06;
Matches 48; Conservative 37; Mismatches 83; Indels 34; Gaps 5;

Qy 45 GMLGFTGYKGRKISVMGHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDV 104
Db 30 GALFISGSLGNHBEIVVRSGIGKVLASITTSLLIQYGVNMVINTGSAGIGSLQVGDI 89
Qy 105 VIGMGACTDSKVNRIKPDHFAA-----ADFDVMRNVADAAKALGIDA 149
Db 90 VIS-----DKVAVFDADATGFGYNPGQLPCMPLYYEASTYLRSEMVKAATNLNA 140
Qy 150 RVGNLFSADLFYSPDGEFDMVEKY-GILGVMEAAAGIYGVAAEFGNAKALTICTVSDHIR 208
Db 141 KEGLIVTGDFTVDSPPKIKEILTNPPEALACEMEAGAAVGQTARQFNIPFLIVRAMSD--- 197
Qy 209 THEQTTAAERQT-TFNDMIKIA 229
Db 198 -----TADHSATQSDFEFEDA 214

RESULT 14
US-10-282-122A-60181
; Sequence 60181, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
```



```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60181
; LENGTH: 232
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-60181

Query Match      11.1%; Score 136; DB 15; Length 232;
Best Local Similarity 23.0%; Pred. No. 2.5e-06;
Matches 53; Conservative 39; Mismatches 102; Indels 36; Gaps 6;

Qy 32 TFLDGA---REVNVVRGMLGFTGYKGRKISVMGHGVGIPSCSIYTKELIITDFGVKKIIR 88
Db 14 TLLRDKIENRQTITIGSGSEIYTGQHGVDVALLKSGIGKVAAMGATILLERCPDVIIN 73
Qy 89 VSGCGAVLPHVKLRDVIIVGMACTDSKVNIRFKDHPAAI-----ADPFD 133
Db 74 TGSAGGLASTLKVGDIVVS-----DEARYHDADVTAFGYEGQLPGCPAGFAKDEK 124
Qy 134 MYRNAVDAAKALGIDARVGNLFSADLFVSPDGEFDMVEKY-GILGVEMEAAGIYGVAAE 192
Db 125 LVAAAESICIKDLNNAVRLIVSGDAFINGSVGLAKIRHNPQAIIVEMEAATAIAHVCHN 184
Qy 193 FGAKALITCTVSD-----HIRTHTQTTAABRQTTFNMDIKIALESVLLG 236
Db 185 FKVPFVVVRAISDVADQOQSHLSPEEFLAARQSTL--MVENLVQNLRG 232

Search completed: July 18, 2005, 21:53:21
Job time : 157 secs
```

```
RESULT 15
US-10-282-122A-55749
; Sequence 55749, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haseelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55749
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 22, 2005, 22:31:58 ; Search time 5335 Seconds
(without alignments)
2170.721 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225
Sequence: 1 MATHPHINAEAGDFADVVLMP.....TTFNDMIKIALESVLLGDKE 239

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool_p/US10035300/runat_18072005.165350.27973/app_query.fasta_1.391
-DB=GenEmbl -QFMT=fastcap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spect -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10035300.cGCGN 1.1 5600 @runat_18072005.165350.27973 -NCPU=6 -ICPU=3
-NO MMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.ats.*
12: gb.sv.*
13: gb.un.*
14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1222	99.8	720	6	BD187724	BD187724 A method
2	1222	99.8	720	6	CQ774461	CQ774461 Sequence
3	1222	99.8	720	6	AX087917	AX087917 Sequence
4	1222	99.8	720	6	AX590437	AX590437 Sequence

5	1222	99.8	1249	1	ECOPNP	M60917 E.coli puri
6	1222	99.8	3031	6	BD261823	BD261823 Recombina
7	1222	99.8	3031	6	AX027820	AX027820 Sequence
8	1222	99.8	3128	6	BD261824	BD261824 Recombina
9	1222	99.8	3128	6	AX027821	AX027821 Sequence
10	1222	99.8	3383	6	BD261814	BD261814 Recombina
11	1222	99.8	3383	6	AX027811	AX027811 Sequence
12	1222	99.8	3934	6	BD261825	BD261825 Recombina
13	1222	99.8	3934	6	AX027822	AX027822 Sequence
14	1222	99.8	4189	6	BD261816	BD261816 Recombina
15	1222	99.8	4189	6	AX027813	AX027813 Sequence
16	1222	99.8	5013	6	AR264513	AR264513 Sequence
17	1222	99.8	5241	6	BD261818	BD261818 Recombina
18	1222	99.8	5241	6	AX027815	AX027815 Sequence
19	1222	99.8	5495	6	BD261815	BD261815 Recombina
20	1222	99.8	5495	6	AX027812	AX027812 Sequence
21	1222	99.8	6046	6	BD261826	BD261826 Recombina
22	1222	99.8	6046	6	AX027823	AX027823 Sequence
23	1222	99.8	6269	6	BD261820	BD261820 Recombina
24	1222	99.8	6269	6	AX027817	AX027817 Sequence
25	1222	99.8	6299	6	BD261821	BD261821 Recombina
26	1222	99.8	6299	6	AX027818	AX027818 Sequence
27	1222	99.8	6301	6	BD261817	BD261817 Recombina
28	1222	99.8	6301	6	AX027814	AX027814 Sequence
29	1222	99.8	11599	1	AE005669	AE005669 Escherich
30	1222	99.8	39675	1	U00096_46	Continuation (47 o
31	1222	99.8	110000	2	AC074221_1	Continuation (2 of
32	1222	99.8	168764	1	AP002569	AP002569 Escherich
33	1222	99.8	338534	1	ECOUW93	U14003 Escherichia
34	1221	99.7	10535	1	AE015447	AE015447 Shigella
35	1221	99.7	86988	1	AE016772	AE016772 Escherich
36	1221	99.7	225944	1	AE016993	AE016993 Shigella
37	1213	99.0	720	6	AX087929	AX087929 Sequence
38	1187	96.9	88037	1	AL627284	AL627284 Salmonell
39	1187	96.9	284233	1	AE016849	AE016849 Salmonell
40	1182	96.5	21405	1	AE008915	AE008915 Salmonell
41	1162	94.9	732	6	AR384902	AR384902 Sequence
42	1115	91.0	110000	1	EX936398_06	Continuation (7 of
43	1106	90.3	11093	1	AE013977	AE013977 Versinia
44	1106	90.3	214050	1	AJ414142	AJ414142 Versinia
45	1106	90.3	293371	1	AE017141	AE017141 Versinia

ALIGNMENTS

BD187724	BD187724	720 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	A method of producing a cytosine nucleoside compound.				
DEFINITION	A method of producing a cytosine nucleoside compound.				
ACCESSION	BD187724.1 GI:32997463				
VERSION	BD187724.1 GI:32997463				
KEYWORDS	JP 2003018997-A/3.				
SOURCE	Escherichia coli				
ORGANISM	Escherichia coli				
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.				
AUTHORS	Akaki, T., Ikeda, I., Matsuishi, K., Abe, R., Oikawa, T., Matsuba, Y., Nagahara, K., Fukui, Y. and Ishibashi, H.				
TITLE	A method of producing a cytosine nucleoside compound				
JOURNAL	MITSUI CHEMICALS INC				
COMMENT	OS Escherichia coli PN JP 2003018997-A/3 PD 21-JAN-2003 PF 01-MAY-2002 JP 2002129867 PI TADASHI ARAKI, ICHIRO IKEDA, KAORI MATOISHI, REIKO ABE, TOSHIHIRO OIKAWA, PI YASUKO MATSUBA, KIYOTERU NAGAHARA, YASUSHI FUKUI, HIROKI ISHIBASHI PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/10 PC C12P19/40, C12N15/00, PC C12N5/00				

CC A method of producing a cytosine nucleoside compound FH Key
FT source Location/Qualifiers
FT source 1..720 /organism="Escherichia coli".

FEATURES
source Location/Qualifiers
1..720 /organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"

ORIGIN

Alignment Scores:
Pred. No.: 4.06e-110 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservativeness: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x BD187724 (1-720)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTTCGCTGACGTAGTTTTCATGCCA 60
Qy 21 GlyAspProLeuArgAlaIleValThrPheLeuGluAspAlaArgGluVal 40
Db 61 GCGACCCGCTGCGTGGGAGATATTGCTGAAACTTTCCTTGAAGATGCCGCTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerVal 60
Db 121 AACACGTTCCGCGTATGCTGGGCTTACCCGCTTACAAAGCGCCAAATTTTCGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleThrLysGluLeuIleThrAsp 80
Db 181 ATGGCTCACGCTATGGGTATCCCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
Qy 81 PheGlyValIleValIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGCGTGAAGAAATTTATCCGCTGGGTTCCTGTCGCGAGTTCTGCGCGACGTAAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGCGACGCTGCTTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAACCGCATCG 360
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGCGCTATCGCTGAGTTCGCTGAGTTCCTCCGCTGACCTGTT 480
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGACACTGGGTATTTGATGCTGGGTAACTTCTCCGCTGACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACGCGGAAATGTTCCGCTGATGAGAAATACGSCATTTCTCGGCGTGAA 540
Qy 181 MetGluAlaAlaGlyIleThrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAACGCGCTGGTATCTACGCGCTGCTGCAGAAATTTGCGCGAAAGCCCTGACCAT 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCCGACATCACAGGACGACCATGCGCTGAGCGTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGGATAAAGAG 717

RESULT 2

CQ774461
LOCUS 720 bp DNA linear PAT 06-MAR-2004
DEFINITION Sequence 23 from Patent WO2004013333.

ACCESSION CQ774461
VERSION CQ774461.1 GI:45237695
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Kock, M., Frank, M. and Badur, R.
TITLE Novel selection method
JOURNAL Patent: WO 200401333-A 23 12-FEB-2004;
BASIS BASF Plant Science GmbH (DE)
FEATURES Location/Qualifiers
source 1..720
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
CDS 1..720
/notes="unnamed protein product; coding for purine nucleoside phosphorylase (deod)"
/codon_start=1
/transl_table=11
/protein_id="CAF34364.1"
/db_xref="GI:45237695"
/translation="MATPHINAEEMGDADVLMFGDPLRAKYIAETFLDAREVNNVR
GMLFGTGYTKRKISVMHGMIIPSCSIYTKELITDFGVKKIIRVSCGVALPHVKLR
DVIVNGACTDSKVNIIRFKDHPAIAIDPDMVNAVDAAKALGIDARVGNLFSADLF
YSPDGMFDMEKYIGILGVMEAAAGIYVNAEFGAKALITCTVSDHIRTHEQTAAER
QTTFNDMIKIALESVLLGDKE"

ORIGIN

Alignment Scores:
Pred. No.: 4.06e-110 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservativeness: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x CQ774461 (1-720)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTTCGCTGACGTAGTTTTCATGCCA 60
Qy 21 GlyAspProLeuArgAlaIleValThrPheLeuGluAspAlaArgGluVal 40
Db 61 GCGACCCGCTGCGTGGGAGATATTGCTGAAACTTTCTTGAAGATGCCGCTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerVal 60
Db 121 AACACGTTCCGCGTATGCTGGGCTTACCCGCTTACAAAGCGCCAAATTTTCGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleThrLysGluLeuIleThrAsp 80
Db 181 ATGGCTCACGCTATGGGTATCCCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
Qy 81 PheGlyValIleValIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGCGTGAAGAAATTTATCCGCTGGGTTCCTGTCGCGAGTTCTGCGCGACGTAAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGCGACGCTGCTTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAACCGCATCG 360
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGCGCTATCGCTGAGTTCGCTGAGTTCCTCCGCTGACCTGTT 480
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGACACTGGGTATTTGATGCTGGGTAACTTCTCCGCTGACCTGTTTC 480

```
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTTCGGGACGGCGAAATGTTCCAGCTGATGGGAAATACGGCATTCCTCGGCGTGAA 540
Qy 181 MetGluAlaGlyIleTyrGlyValAlaGluPheGlyAlaValAlaLeuThrIle 200
Db 541 ATGGAGCGGCTGGTATCTACGGCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCGCACTCAGGACAGACCACTCGCGCTGAGCGTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 717

RESULT 3
AX087917 LOCUS AX087917 720 bp DNA linear PAT 17-MAR-2001
DEFINITION Sequence 3 from Patent WO0114566.
ACCESSION AX087917
VERSION AX087917.1 GI:13396895
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
AUTHORS Tischer, W., Ihlenfeldt, H.G., Barzu, O., Sakamoto, H., Pistotnik, E.,
Marliere, P., and Pochet, S.
TITLE Enzymatic synthesis of deoxyribonucleosides
JOURNAL Patent: WO 0114566-A 3 01-MAR-2001;
Roche Diagnostics GmbH (DE); INSTITUT PASTEUR (FR); Pharma-
Waldhof GmbH & Co. KG (DE)
FEATURES
source Location/Qualifiers
1..720
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
CDS
1..720
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC34665.1"
/db_xref="GI:13396896"
/transl_table="1"
/translation="MATPHINAEAGMDPADVVLMPGDLRAKVIATFLEDAEYNNVR
GMLGPTGYKGRKISVMGHGNGIPSCSIYTKELJTDGKVKIIRVSGCGLVPHVKLR
DVVIGMGACTDSKVNRIKPDHFAALIDFMRNADVAALGIDARVGNLFSADLF
YSPGEMPDVMEKYGILGVEVEAAGIYGVAABFGAKALITICTVSDHIRTHTQTTAAER
QTTFNMIKIALSVLLIGDKE"
ORIGIN
Alignment Scores:
Pred. No.: 4.06e-110 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0
US-10-035-300A-2 (1-239) x AX087917 (1-720)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGCTACCCACACATTAATGACAGAAATGGCGAATTCGCTGAGTAGTTTGTATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GCGACCGCGCTGCGTGGCAAGTATATTCCTGAAGATGCCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyValArgLysIleSerVal 60
Db 121 AACAAAGCTTCGCGGTATGCTGGGCTTACCGGTAATCTTACAAAGGCCCGCAAAATTTCCGTA 180
```

```
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 181 ATGGGTACGGGTATGGGTATCCCGTCTCTCTACCAAGAACTGATCACCGAT 240
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGGCGTGAAGAAATATATCCCGTGGGTTCCTGTGGCGAGTTCTCCGCGACGTAATA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGGCGGACGTCGTATCGGTATGGGTGCTGCACCGATTCGAAAGTTTAAACCGCATCCGT 360
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGCGCTATCGCTGACTTCGACATGGTGGCTAAACGACGTAGAT 420
Qy 141 AlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGCACATGGGTATTTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCGGACGGCGAAATGTTTCAGCTGATGGAATAATACGGCATTCCTCGCGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAAGCGGCTGCTATCTACGGCGTCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCGCACTCAGGACAGACCACTCGCGCTGAGCGTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 717

RESULT 4
AX590437 LOCUS AX590437 720 bp DNA linear PAT 27-JAN-2003
DEFINITION Sequence 3 from Patent EP1254959.
ACCESSION AX590437
VERSION AX590437.1 GI:27949070
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Escherichia coli
REFERENCE 1 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
AUTHORS Araki, T., Ikeda, I., Matoishi, K., Abe, R., Oikawa, T., Matsuba, Y.,
Ishibashi, H., Nagahara, K. and Fukui, Y.
TITLE Method for producing cytosine nucleoside compounds
JOURNAL Patent: EP 1254959-A 3 06-NOV-2002;
MITSUI CHEMICALS, INC. (JP)
FEATURES
source Location/Qualifiers
1..720
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
ORIGIN
Alignment Scores:
Pred. No.: 4.06e-110 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0
US-10-035-300A-2 (1-239) x AX590437 (1-720)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
```

```
Dbb 1 ATGGCTACCCACACATTAATGACAGAAATGGCGATTTTCGCTGACGTAGTTTTGTATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Dbb 61 GGGACCCGCTGCGTGGGAAGTATTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Dbb 121 AACACGTTCCGGGTATGCTGGGCTTACCGGTACTTACAAAGCCGCAAAATTCGGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Dbb 181 ATGGGTCAAGTATGGGTATCCGCTGCTGCTCCATCTACACCAAGAACTGATCACCAGT 240
Qy 81 PheGlyValLysIleLysIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Dbb 241 TTGGCGGTGAAGAAATATTCCGCGTGGGTTCCTGTGGCGCAGTTCTGCCGACGTAATA 300
Qy 101 LeuArgAspValValIleGlyMetGlyValaCysThrAspSerLysValAsnArgIleArg 120
Dbb 301 CTGGCGACGTCGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 360
Qy 121 PheLysAspHisAspPheAlaLysIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Dbb 361 TTTAAAGACCATGACTTTGCGGTATCGCTGACTTCGACATGGTGGCTAAGCAGTAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Dbb 421 GCAGCTAAAGACACTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTT 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Dbb 481 TACTCTCCGCGCGGAAATGTTCCAGCTGATGGAATAATACGGCATTTCTCGCGGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Dbb 541 ATGGAAGCGGCTGGTATCTACGCGGTGCTGCAAGATTTGCGCGGAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Dbb 601 TGACCGTATCTGACCACTACCGGCTGCTGCAAGATTTGCGCGGAAGCCCTGACCATC 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Dbb 661 ACCTTCAACACATGATCAAAATCGCACTCGAATCCGTTCTGCTGGCGGATAAAGAG 717
```

```
RESULT 5
ECOPNP 1249 bp DNA linear BCT 26-APR-1993
LOCUS E.coli purine nucleoside phosphorylase (deod) gene, complete cds.
DEFINITION M60917
ACCESSION M60917
VERSION M60917.1 GI:147308
KEYWORDS purine nucleoside phosphorylase.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 1249)
AUTHORS Hershfield,M.S., Chaffee,S., Koro-Johnson,L., Mary,A., Smith,A.A.
and Short,S.A.
TITLE Use of site-directed mutagenesis to enhance the epitope-shielding
effect of covalent modification of proteins with polyethylene
glycol
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88 (16), 7185-7189 (1991)
MEDLINE 91334430
PUBMED 1714590
COMMENT source text: Escherichia coli (strain K-12) DNA.
FEATURES
Original Location/Qualifiers
1..1249
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/db_xref="taxon:562"
```

```
RBS 111..115
/notes="putative"
gene 123..842
/genes="deod"
CDS 123..842
/genes="deod"
/EC_number="2.4.2.1"
/notes="putative"
/codon_start=1
/transl_table=11
/product="purine nucleoside phosphorylase"
/protein_id="AAA24401.1"
/db_xref="GI:147308"
translation="MATPHINAEIMGDFADVLMPGDPDLRAKYIAETFLBDAREVNVNR
GMVFTGTGKRIKISVMGHMGIPSCSIYTKELITDPGVKKIIRVSCGAVLPHVKLR
DVVIGMGACTDSKNRIRFADHDFAAIADFDMVRNVAADAAKALGIDARVGNLFSADLF
YSPDGMFDMVMEKYIGILGVEMEAAGIYVAAEFGAKALITICTVSDHRTHEQTAAER
QTTFNMIKIALESVLLGDKE"
stem_loop 874..901
/notes="putative"
/function="terminator for deod transcription"
ORIGIN
Alignment Scores:
Pred. No.: 8,17e-110 Length: 1249
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: Gaps: 0
US-10-035-300A-2 (1-239) x ECOPNP (1-1249)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Dbb 123 ATGGCTACCCACACATTAATGCAAAATGGCGATTTTCGCTGACGTAGTTTTGTATGCCA 182
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Dbb 183 GGGACCCGCTGCGTGGGAAGTATTGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 242
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Dbb 243 AACACGTTCCGGGTATGCTGGGCTTCAACCGGTACTTACAAAGCGCGCAAAATTCGTA 302
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Dbb 303 ATGGGTCAAGGTATGGGTATCCGCTCCTGCTCATCTACACCAAGAACTGATCACCAGT 362
Qy 81 PheGlyValLysIleLysIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Dbb 363 TTGGCGGTGAAGAAATTTATCCGCTGGGTTCCTGTGGCGCAGTTCTGCCGACGTAAAA 422
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Dbb 423 CTGCGGACGCTGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 482
Qy 121 PheLysAspHisAspPheAlaLysIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Dbb 483 TTTAAAGACCATGACTTTCCCGCTATCGGTGACTTCGACATGGTGGGTAAACGAGTAGAT 542
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Dbb 543 GCAGCTAAAGCACTGGGTATTGATGCTCGGTGGGTAACTGTTCTCCGCTGACCTGTTT 602
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Dbb 603 TACTCTCCGACGCGGAAATGTTCCGACGTGATGGAATAATACGGCATTTCTCGCGGTGAA 662
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Dbb 663 ATGGAAGCGGCTGGTATCTACCGGCTGCTGCAAGATTTGGCGCGGAAGCCCTGACCATC 722
```

```

Qy 201 CysThrValSerAspHisIleAArgThrHisGluInThrThrAlaAlaGluArgGlnThr 220
Db 723 TGCACCGTATCTGACCATCGCACTCACGAGCAGACCACTCGCGCTGAGCGTCAGACT 782

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGluAspLysGlu 239
Db 783 ACCTTCAACGACATGATCAAAATGCACTGGATCCGATTCCTGCTGGCGCATAAAGAG 839

RESULT 6
BD261823 3031 bp DNA linear PAT 17-JUL-2003
LOCUS Recombinant bacterial strains for the production of natural
DEFINITION nucleosides and modified analogues thereof.
ACCESSION BD261823.1 GI:33071591
VERSION JP 2002533126-A/12.
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 3031)
AUTHORS Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL Patent: JP 2002533126-A 12 08-OCT-2002;
NORPHARMA SPA
COMMENT OS Artificial Sequence
PN JP 2002533126-A/12
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI, SIMONA CALI, DANIELA GHISOTTI, GAETANO PI
ORSINI,
PI GIANCARLO TONON, GABRIELE ZUFFI
PC C12N15/09, C12N1/21, C12N9/10//C12P19/38, C12P19/40, C12N15/00 CC
Description of Artificial Sequence: udp and deod cloned into CC
PGW746
CC without upstream ptac promoter
FH Key Location/Qualifiers
FT source 1..3031
FT /organism='Artificial Sequence'.

FEATURES
source
1..3031
Location/Qualifiers
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN
Alignment Scores: 2.52e-109 Length: 3031
Pred. No.: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x BD261823 (1-3031)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 101 ATGGCTACCCACACATTAATGCGAATAATGGCGAATTCGCTGACGTAGTTTGTATGCCA 160

Qy 21 GlyAspProLeuArgAlaLysTyrlleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 161 GGCGACCGCTGCTGGCGAAGTATATGCTCAAACTTCTTGAAGATGCCCGTGAAGTG 220

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerVal 60
Db 221 AACAAACGTTCGCGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 280

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGluLeuIleThrAsp 80
Db 281 ATGGGTACCGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGT 340

```

```

Qy 81 PheGlyValLysLysIleAArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 341 TTGCGCGTGAAGAAAATTTATCCGCGTGGGTTCTCTGTGCGCAGATTCTCGCGCACGTAAA 400

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 401 CTGGCGACGCTGTTATCGGTATGGGTGCTTGCACCGAATTCAAAGTTTAAACCGCATCCGT 460

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 461 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGCGTAACGACGTAGAT 520

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 521 GCAGCTAAAGCACTGGGTATTTGATGCTGCGGTGGGTAACTGTCTTCCGCTGACCTGTC 580

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyIleLeuGlyValGlu 180
Db 581 TACTCTCGGACGGCGGAATTTTCAGCTGATGGAAAATACGGCATTTCTCGGCGTGGAA 640

Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 641 ATGGAAGCGGCTGTTATCTACGCGCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 700

Qy 201 CysThrValSerAspHisIleAArgThrHisGluInThrThrAlaAlaGluArgGlnThr 220
Db 701 TGCACCGTATCTGACCATCATCGCACTCACGAGCAGACCACTGCGGCTGAGCGTCAGAGT 760

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 761 ACCTTCAACGACATGATCAAAATCGCACTGGATCCGTTCTGCTGGCGCATAAAGAG 817

RESULT 7
AX027820 3031 bp DNA linear PAT 16-SEP-2000
LOCUS Sequence 12 from Patent WO0039307.
DEFINITION AX027820
ACCESSION AX027820.1 GI:10188664
VERSION AX027820.1
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL Patent: WO 0039307-A 12 06-JUL-2000;
BESTETTI GIUSEPPINA (IT); CALI SIMONA (IT); NORPHARMA SPA (IT);
ORSINI GAETANO (IT); TONON GIANCARLO (IT); ZUFFI GABRIELE (IT);
GHISOTTI DANIELA (IT)
FEATURES
source
1..3031
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="udp and deod cloned into pGW746, without upstream
ptac promoter"

ORIGIN
Alignment Scores: 2.52e-109 Length: 3031
Pred. No.: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x AX027820 (1-3031)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 101 ATGGCTACCCACACATTAATGCGAATAATGGCGAATTCGCTGACGTAGTTTGTATGCCA 160

```

```
QY 21 GlyAspProLeuArgAlaIysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
|
|
|
Db 161 GGCACCCGCTGGTGGAGATATATGCTGAACCTTCTTGAAGATGCCGTGAAGTG 220
|
|
|
QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrIysGlyArgIysIleSerVal 60
|
|
|
Db 221 AACAAACGCTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCCAAATTTCCGTA 280
|
|
|
QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrIysGluLeuIleThrAsp 80
|
|
|
Db 281 ATGGGTACACGGTATGGGTATCCCGTCTGCTGCATCTACACCAAGAACTGATCACCAGT 340
|
|
|
QY 81 PheGlyValIysIysIleIleArgValGlySerCysGlyAlaValLeuProHisValIys 100
|
|
|
Db 341 TTCCGCGTGAAGAAATATATCCCGGTGGGTCTCTGTGGCGCAGTTCTGCCGACGTAAAA 400
|
|
|
QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerIysValAsnArgIleArg 120
|
|
|
Db 401 CTGCGCGACGCTGTTATCGGTATGGGTGCTGCACCGGATTCCAAAGTTAAACCGCATCCGT 460
|
|
|
QY 121 PheIysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
|
|
|
Db 461 TTTAAAGACCATGACTTTGGCGGTATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTT 520
|
|
|
QY 141 AlaAlaIysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
|
|
|
Db 521 GCAGCTAAAGCATGGGTATGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTTT 580
|
|
|
QY 161 TyrSerProAspGlyGluMetPheAspValMetGluIysTyrGlyIleLeuGlyValGlu 180
|
|
|
Db 581 TACTCTCCGACGCGGAAATGTTCCAGCTGATGGAAATACGCGCATTCCTCGCGGTGGAA 640
|
|
|
QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaIysAlaLeuThrIle 200
|
|
|
Db 641 ATGGAACGGCTGATGATCTACGGCGTGCCTGCGAGATTTGCGCGAAGCCCTGACCATC 700
|
|
|
QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
|
|
|
Db 701 TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCATGCGCGTGAAGCTGAGACT 760
|
|
|
QY 221 ThrPheAsnAspMetIleIysIleAlaLeuGluSerValLeuLeuGlyAspIysGlu 239
|
|
|
Db 761 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATAAGAG 817
|
|
|
RESULT 8
BD261824 3128 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION
Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof.
ACCESSION
BD261824
VERSION
BD261824.1 GI:33071592
KEYWORDS
JP 2002533126-A/13.
SOURCE
synthetic construct
ORGANISM
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 3128)
AUTHORS
Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE
Recombinant bacterial strains for the production of natural
nucleosides and modified analogues thereof
JOURNAL
Patent: JP 2002533126-A 13 08-OCT-2002;
NORPHARMA SPA
COMMENT
OS Artificial Sequence
PN JP 2002533126-A/13
PD 08-OCT-2002
PF 23-DEC-1999 JP 2000591198
PR 23-DEC-1998 IT MI 98A002792
PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI
ORSINI,
PI GIANCARLO TONON,GABRIELE ZUFFI
PC C12N15/09 C12N1/21, C12N9/10/C12P19/38, C12P19/40, C12N15/00 CC
Description of Artificial Sequence: deod cloned downstream CC
ptac promoter
PH Key Location/Qualifiers
```

```
FT source 1..3128 /organism='Artificial Sequence'.
FT Location/Qualifiers
FEATURES
source 1..3128
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Alignment Scores: 2.62e-109 Length: 3128
Pred. No.: 1222.00 Matches: 238
Score: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0
US-10-035-300A-2 (1-239) x BD261824 (1-3128)
QY 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
|
|
|
Db 198 ATGGCTTACCCACACATTAATGCGAAATGGGCGATTCGCTGACGTAGTTTGTATGCCA 257
|
|
|
QY 21 GlyAspProLeuArgAlaIysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
|
|
|
Db 258 GGCACCCGCTGGTGGAGTATATGCTGAACCTTTCCTTGAAGATGCCCGTGAAGTG 317
|
|
|
QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrIysGlyArgIysIleSerVal 60
|
|
|
Db 318 AACAAACGCTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCCAAATTTCCGTA 377
|
|
|
QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrIysGluLeuIleThrAsp 80
|
|
|
Db 378 ATGGGTACACGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCAGT 437
|
|
|
QY 81 PheGlyValIysIysIleIleArgValGlySerCysGlyAlaValLeuProHisValIys 100
|
|
|
Db 438 TTCCGCGGTGAAGAAATATATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGACGTAAAA 497
|
|
|
QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerIysValAsnArgIleArg 120
|
|
|
Db 498 CTGCGGACGCTGTTATCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 557
|
|
|
QY 121 PheIysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
|
|
|
Db 558 TTTAAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGTTGCGTAACGCAGTAGAT 617
|
|
|
QY 141 AlaAlaIysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
|
|
|
Db 618 GCAGCTAAAGCACTGGGTATGATGCTCGCGTGGTAAACCTGTTCTCCGCTGACCTGTTT 677
|
|
|
QY 161 TyrSerProAspGlyGluMetPheAspValMetGluIysTyrGlyIleLeuGlyValGlu 180
|
|
|
Db 678 TACTCTCCGACGCGGAAATGTTGACGCTGATGGAATAATACGCGATTCGCGCGTGGAA 737
|
|
|
QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaIysAlaLeuThrIle 200
|
|
|
Db 738 ATGGAAGCGCTGGTATCTACGGCGTCTGCGAGATTTGCGCGAAGCCCTGACCATC 797
|
|
|
QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
|
|
|
Db 798 TGCACCGTATCTGACCATCCGCACTCAGCAGCAGACCATGCGCGTGAAGCTGAGCGTCACT 857
|
|
|
QY 221 ThrPheAsnAspMetIleIysIleAlaLeuGluSerValLeuLeuGlyAspIysGlu 239
|
|
|
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAGAG 914
|
|
|
RESULT 9
AX027821 3128 bp DNA linear PAT 16-SEP-2000
LOCUS
DEFINITION
Sequence 13 from Patent WO0039307.
ACCESSION
AX027821
VERSION
AX027821.1 GI:10188665
```



```
KEYWORDS      .
SOURCE         synthetic construct
ORGANISM       other sequences; artificial sequences.
REFERENCE      1
AUTHORS        Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE          Recombinant bacterial strains for the production of natural
                nucleosides and modified analogues thereof
JOURNAL        Patent: WO 0039307-A 13 06-JUL-2000;
                BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ;
                ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
                GHISOTTI DANIELA (IT)
FEATURES       Location/Qualifiers
                source
                1..3128
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="deod cloned downstream ptac promoter"
ORIGIN
Alignment Scores:      2.62e-109      Length:      3128
Pred. No.:            1222.00      Matches:      238
Percent Similarity:    100.00%      Conservative: 1
Best Local Similarity: 99.58%      Mismatches:   0
Query Match:          99.76%      Indels:       0
DB:                   6           Gaps:         0
US-10-035-300A-2 (1-239) x AX027821 (1-3128)
Qy      1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db      198 ATGGCTACCCACACATTAATGCAGAAATGGCGCATTCGCTGACGTAGTTTGTATGCCA 257
Qy      21 GlyAspProLeuArgAlaLysTyriLeAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db      258 GGCACCGCGTCGCTGGAGAGTATTGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 317
Qy      41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyriLysGlyArgLysIleSerVal 60
Db      318 AACAACTTCGCGGTATGCTGGCTTCACCGTACTTACAAAGCGCGCAAAATTTCCGTA 377
Qy      61 MetGlyHisGlyValGlyIleProSerCysSerIleTyriThrLysGluLeuIleThrAsp 80
Db      378 ATGGGTACCGGTATGGGTATGCCGCTCTGCTCCATCTACACAAAGAACTGATCACCAGT 437
Qy      81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db      438 TTCGCGGTGAAGAAATATATCCGCGTGGGTCTCTGTGGCGAGTTCGCCGACGTAAAA 497
Qy      101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db      498 CTGCGCAGCTCGTTATCGGTATGGGTATGGTGCCTGCACCGATTCCAAAGTTAACCGCATCCGT 557
Qy      121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db      558 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGCCTGAACGAGTAGAT 617
Qy      141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db      618 GCACCTAAGACACTGGGTATTGATGCTCGCGTGGGTAACTTCTCCGCTGACCTGTTC 677
Qy      161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyriGlyIleLeuGlyValGlu 180
Db      678 TACTCTCCGACGCGGAAATGTTCCGACGTATGTAAGAAATAACGGCATTCCTCGCGGTGAA 737
Qy      181 MetGluAlaAlaGlyIleTyriGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db      738 ATGGAAGCGGTGTTATCTACGGGTCTGCTGAGAAATTTGGCGGAAAGCCCTGACCATC 797
Qy      201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db      798 TGCACCGGTATCTGACCATATCCGCACTCAGCAGCAGACCACTGCCGCTGACCGTCAGACT 857
Qy      221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db      858 ACCTTCAACGACATGATCAAAATCGCAATCCGTTCTGCTGGCGGATAAAGAG 914
RESULT 10
BD261814      3383 bp      DNA      linear      PAT 17-JUL-2003
LOCUS         Recombinant bacterial strains for the production of natural
DEFINITION    nucleosides and modified analogues thereof.
ACCESSION     BD261814
VERSION       BD261814.1 GI:33071582
KEYWORDS      JP 2002533126-A/3.
SOURCE        synthetic construct
ORGANISM      other sequences; artificial sequences.
REFERENCE      1 (bases 1 to 3383)
AUTHORS       Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE         Recombinant bacterial strains for the production of natural
                nucleosides and modified analogues thereof
JOURNAL        Patent: JP 2002533126-A 3 08-OCT-2002;
                NORPHARMA SPA
COMMENT        OS Artificial Sequence
                PN JP 2002533126-A/3
                PD 08-OCT-2002
                PF 23-DEC-1999 JP 2000591198
                PR 23-DEC-1998 IT MI 98A002792
                PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI
                ORSINI,
                PI GIANCARLO TONON GABRIELE ZUFFI
                PC C12N15/09,C12N1/21,C12N9/10//C12P19/38,C12P19/40,C12N15/00 CC
                Description of Artificial Sequence: Plasmid
                CC deod
                FH Key      Location/Qualifiers
                FT gene      (231)..(960).
                source
                1..3383
                /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
ORIGIN
Alignment Scores:      2.9e-109      Length:      3383
Pred. No.:            1222.00      Matches:      238
Percent Similarity:    100.00%      Conservative: 1
Best Local Similarity: 99.58%      Mismatches:   0
Query Match:          99.76%      Indels:       0
DB:                   6           Gaps:         0
US-10-035-300A-2 (1-239) x BD261814 (1-3383)
Qy      1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db      240 ATGGCTACCCACACATTAATGCAGAAATGGCGCATTCGCTGACGTAGTTTGTATGCCA 299
Qy      21 GlyAspProLeuArgAlaLysTyriLeAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db      300 GGCACCGCGTCGCTGGAGAGTATTGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 359
Qy      41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyriLysGlyArgLysIleSerVal 60
Db      360 AACAACTTCGCGGTATGGGTATGGGTCTTCCCGTATCTACAAAGCGCGCAAAATTTCCGTA 419
Qy      61 MetGlyHisGlyValGlyIleProSerCysSerIleTyriThrLysGluLeuIleThrAsp 80
Db      420 ATGGGTACCGGTATGGGTATCCCGTCTCTCCATCTACACAAAGAACTGATCACCAGT 479
Qy      81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db      480 TTCGCGCTGAAGAAATATATCCGCGTGGGTTCCTGTGGCGAGTTCGCCGACGTAAAA 539
Qy      101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
```

```
Db      540  CTGCGGACGTCGTATCGGTATGGGTGCGCTGCACCGATTCCAAAGTTAACCGCATCCGT 599
Qy      121  PheLyAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db      600  TTTTAAAGACCATGACTTTGCGCGTATCGCTGACTTCGACATGGTGGCTAACGCGTAGAT 659
Qy      141  AlaAlaLyAlaLeuGlyIleAspAlaAatqValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db      660  GCAGCTAAAGCACTGGGTATGTATCGGTGGGTAACTGTCTCCGCTGACCTGTTTC 719
Qy      161  TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db      720  TACTCTCCGACGCGGAAATGTTCCGACGTATCGGATGCGGATTCCTCGCGGTGAA 779
Qy      181  MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db      780  ATGGAAGCGGCTGGTATCTACGGCGTGGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 839
Qy      201  CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db      840  TGCACCGTATCTGACCACATCCGCACTTCAGACGACGACCACTGCCGCTGAGCGTCAGACT 899
Qy      221  ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db      900  ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 956

RESULT 11
AX027811
LOCUS      AX027811          3383 bp      DNA      linear      PAT 16-SEP-2000
DEFINITION Sequence 3 from Patent WO0039307.
ACCESSION  AX027811
VERSION     AX027811.1  GI:10188655
KEYWORDS   .
SOURCE      synthetic construct
            other sequences; artificial sequences.
ORGANISM    1
REFERENCE   1
AUTHORS     Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE       Recombinant bacterial strains for the production of natural
            nucleosides and modified analogues thereof
JOURNAL     Patent: WO 0039307-A 3 06-JUL-2000;
            BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ;
            ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
            GHISOTTI DANIELA (IT)
FEATURES    Location/Qualifiers
            source          1..3383
                        /organism="synthetic construct"
                        /mol_type="unassigned DNA"
                        /db_xref="taxon:32630"
                        /note="Plasmid"

ORIGIN
Alignment Scores:
Pred. No.:      2..9e-109      Length:      3383
Score:          1222.00      Matches:      238
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 99.58%      Mismatches: 0
Query Match:    99.76%      Indels:      0
DB:             6      Gaps:      0

US-10-035-300A-2 (1-239) x AX027811 (1-3383)

Qy      1  MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db      240  ATGGCTACCCACACATTAATGAGAAATGGCGATTCGCTCAGCTAGTATTTGATGCCA 299
Qy      21  GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db      300  GGGGACCCGCTGGTGGGAGTATATGCTGAACCTTCTCTGAAGATGCCCGTGAAGTG 359
Qy      41  AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
```

```
Db      360  AACACAGTTTCGGGTATGCTGGGCTTTCACCGGTACTTACAAAGGCGCAAAATTTCCGTA 419
Qy      61  MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db      420  ATGGGTACAGGTATGGGTATCCGCTCTGCTCATCTACACCAAGAACTGATCACCAGT 479
Qy      81  PheGlyValLysLysIleIleArgValGlySerCysGlyValaValLeuProHisValLys 100
Db      480  TTCGGGCTGAGNAANATTATCCGCGTGGGTTCCTGTGGCGCAGTTCCTGCCGACGTAAA 539
Qy      101  LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db      540  CTGCGGACGCTGTTATCGGTATGGGTGCGCTGCACCGATTCCAAAGTTAACCGCATCCGT 599
Qy      121  PheLyAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db      600  TTTTAAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGGTGGGTAAACCGTAGAT 659
Qy      141  AlaAlaLyAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db      660  GCAGCTAAAGCACTGGGTATGTATCGGTGGGTAACTGTCTCCGCTGACCTGTTTC 719
Qy      161  TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db      720  TACTCTCCGACGCGGAAATGTTCCGACGTATCGGATGCGGATTCCTCGCGGTGAA 779
Qy      181  MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db      780  ATGGAAGCGGCTGGTATCTACGGCGTGGCTGCAGAAATTTGGCGCGAAAGCCCTGACCATC 839
Qy      201  CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db      840  TGCACCGTATCTGACCACATCCGCACTTCAGACGACGACCACTGCCGCTGAGCGTCAGACT 899
Qy      221  ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db      900  ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 956

RESULT 12
BD261825
LOCUS      BD261825          3934 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Recombinant bacterial strains for the production of natural
            nucleosides and modified analogues thereof.
ACCESSION  BD261825
VERSION     BD261825
KEYWORDS   1  GI:33071593
SOURCE      JP 2002533126-A/14.
            synthetic construct
            other sequences; artificial sequences.
REFERENCE   1  (bases 1 to 3934)
AUTHORS     Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE       Recombinant bacterial strains for the production of natural
            nucleosides and modified analogues thereof
JOURNAL     Patent: JP 2002533126-A 14 08-OCT-2002;
            NORPHARMA SPA
COMMENT     OS Artificial Sequence
            PN JP 2002533126-A/14
            PD 08-OCT-2002
            PF 23-DEC-1999 JP 2000591198
            PR 23-DEC-1998 IT MI 98A002792
            PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI
            ORSINI,
            PI GIANCARLO TONON,GABRIELE ZUFFI
            PC C12N15/09,C12N1/21,C12N9/10//C12P19/38 C12P19/40,C12N15/00 CC
            Description of Artificial Sequence: udp and deod cloned CC
            downstream ptac
            CC promoter
            FH Key      Location/Qualifiers
            FT source   1..3934
            FT          /organism='Artificial Sequence'.
            FT          Location/Qualifiers
            FT          1..3934
            /organism="synthetic construct"
```

```
ORIGIN
/mol_type="genomic DNA"
/db_xref="taxon:32630"

Alignment Scores:
Pred. No.: 3,51e-109 Length: 3934
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x BD261825 (1-3934)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 257
Qy 21 GlyAspProLeuArgAlaLysTyrlleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GCGCACCGCGTGGCGAAGTATATTCGTAACCTTTCTTGAAGATGCCCGTGAAGTG 317
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyValSerVal 60
Db 318 AACAAAGTTCCGGTATTCGCGTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGluLeuIleThrAsp 80
Db 378 ATGGCTACGGTATGGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 437
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGCGGTGAAGAAATATATCCGCGTGGGTTCCTGTGGCGAGTTCTCCCGCACGTAAA 497
Qy 101 LeuArgAspValIleGlyMetGlyValaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTGGCGACGTCGTTATCGGTATGGGTGCTGACCGGATTCCAAAGTTAAACCGCATCCGT 557
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGACTTTGGCGCTATCGCTGACATTCGACATGGTGGTAAACGCAGTAGAT 617
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCAGCTAAAGCACCTGGGTATTTGATGCTCGCGTGGGTAACTTCTCCGCTGACCTGTTTC 677
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyValGlu 180
Db 678 TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAAATACGGCATTTCTCGCGGTGGAA 737
Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGGAAAGCGGTGGTATCTACGGCGTCCGTCAGAAATTTGGCGCAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGGTATCTGACCATCCGCACTCACGAGCAGACCACTCGCGCTGAGCGTCAGACT 857
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 914

RESULT 13
AX027822 3934 bp DNA linear PAT 16-SEP-2000
LOCUS
DEFINITION Sequence 14 from Patent W00039307.
ACCESSION AX027822
VERSION AX027822.1 GI:1018866
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
```

```
AUTHORS Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE Recombinant bacterial strains for the production of natural
JOURNAL nucleosides and modified analogues thereof
Patent: WO 0039307-A 14 06-JUL-2000;
BESTETTI GIUSEPPINA (IT); CALI SIMONA (IT); NORPHARMA SPA (IT);
ORSINI GAETANO (IT); TONON GIANCARLO (IT); ZUFFI GABRIELE (IT);
GHISOTTI DANIELA (IT)
FEATURES
source Location/Qualifiers
1..3934
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="udp and deod cloned downstream ptac promoter"
```

```
ORIGIN
Alignment Scores:
Pred. No.: 3,51e-109 Length: 3934
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 6 Gaps: 0
```

```
US-10-035-300A-2 (1-239) x AX027822 (1-3934)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 257
Qy 21 GlyAspProLeuArgAlaLysTyrlleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GCGCACCGCGTGGCGAAGTATATTCGTAACCTTTCTTGAAGATGCCCGTGAAGTG 317
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyValSerVal 60
Db 318 AACAAAGTTCCGGTATTCGCGTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGluLeuIleThrAsp 80
Db 378 ATGGCTACGGTATGGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 437
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGCGGTGAAGAAATATATCCGCGTGGGTTCCTGTGGCGAGTTCTCCCGCACGTAAA 497
Qy 101 LeuArgAspValIleGlyMetGlyValaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTGGCGACGTCGTTATCGGTATGGGTGCTGACCGGATTCCAAAGTTAAACCGCATCCGT 557
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGACTTTGGCGCTATCGCTGACATTCGACATGGTGGTAAACGCAGTAGAT 617
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCAGCTAAAGCACCTGGGTATTTGATGCTCGCGTGGGTAACTTCTCCGCTGACCTGTTTC 677
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyValGlu 180
Db 678 TACTCTCCGGACGGCGAAATGTTTCGACGTGATGGAAATACGGCATTTCTCGCGGTGGAA 737
Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGGAAAGCGGTGGTATCTACGGCGTCCGTCAGAAATTTGGCGCAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGGTATCTGACCATCCGCACTCACGAGCAGACCACTCGCGCTGAGCGTCAGACT 857
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 914
```

```

RESULT 14
BD261816          4189 bp      DNA      linear      PAT 17-JUL-2003
LOCUS
DEFINITION      Recombinant bacterial strains for the production of natural
                  nucleosides and modified analogues thereof.
ACCESSION      BD261816
VERSION        BD261816.1 GI:33071584
KEYWORDS       JP 2002533126-A/5.
SOURCE         synthetic construct
ORGANISM       other sequences; artificial sequences.
REFERENCE      1 (bases 1 to 4189)
AUTHORS        Bestetti,G., Cali,S., Ghisotti,D., Orsini,G., Tonon,G. and Zuffi,G.
TITLE          Recombinant bacterial strains for the production of natural
                  nucleosides and modified analogues thereof
JOURNAL        Patent: JP 2002533126-A 5 08-OCT-2002;
                  NORPHARMA SPA
COMMENT        OS Artificial Sequence
                  PN JP 2002533126-A/5
                  PD 08-OCT-2002
                  PF 23-DEC-1999 JP 2000591198
                  PR 23-DEC-1998 IT MI 98A002792
                  PI GIUSEPPINA BESTETTI,SIMONA CALI,DANIELA GHISOTTI,GAETANO PI
                     ORSINI,
                     PI GIANCARLO TONON,GABRIELE ZUFFI
                     PC C12N15/09,C12N1/21,C12N9/10/C12P19/38,C12P19/40,C12N15/00 CC
                     Description of Artificial Sequence: Plasmid
                     CC udp
                     CC deod
                     FH Key
                     FT gene
                     FT gene
FEATURES        Location/Qualifiers
source          1..4189
                /organism="synthetic construct"
                /mol_type="genomic DNA"
                /db_xref="taxon:32630"
ORIGIN
Alignment Scores:
Pred. No.:      3.8e-109      Length:      4189
Score:          1222.00      Matches:      238
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 99.58%      Mismatches: 0
Query Match:    99.76%      Indels:      0
DB:             6      Gaps:      0
US-10-035-300A-2 (1-239) x BD261816 (1-4189)
Qy              1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db              1046 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTGCTGACGTAGTTTGTATGCCA 1105
Qy              21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db              1106 GCGCACCGCTGCTGCGAAGTATATTGCTGAACCTTCTCTTGAAGATGCCGGAAGTG 1165
Qy              41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db              1166 AACAAAGCTTCGCGTATGCTGGGCTTCACCGTACTTACAAAGCCGCAAAATTTCCGTA 1225
Qy              61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuLeuThrAsp 80
Db              1226 ATGGGTACCGGTATGGGTATCCCTCTGCTCCATCTACCAAGAACGTGATCACCAGT 1285
Qy              81 PheGlyValLysTyrIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db              1286 TTCGGCTGAAGAAATATATCCGCTGGGTTCCTGTGGCGAGTTCCTGCGCACGTAAAA 1345
Qy              101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db              1346 CTGCGCAGCGTGTATTCGGTATGGGTGCGCTGCACCGATTCCAAAGTTAACCCGATCCGT 1405

```

```

Qy              121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db              1406 TTTAAGACCATGACATTGCGGCTATCGCTGACTTCGACATGTCGTAACGCAGTAGAT 1465
Qy              141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db              1466 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGTAACCTGTTCTCCGCTGACCTGTTTC 1525
Qy              161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db              1526 TACTCTCCGACGCGGAAATGTTGACGCTGATGGAATAATACGGCATTTCTCGCGGTGAA 1585
Qy              181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db              1586 ATGGAAGCGCTGGTATCTACGGCGTCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 1645
Qy              201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db              1646 TGCACCGTATCTGACCACATCCGCACTCACGAGCAGACCACTCCGCTGAGCGTCAGACT 1705
Qy              221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db              1706 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTCTGCTGGCGGATAAGAG 1762
RESULT 15
AX027813          4189 bp      DNA      linear      PAT 16-SEP-2000
LOCUS
DEFINITION      Sequence 5 from Patent WO0039307.
ACCESSION      AX027813
VERSION        AX027813.1 GI:10188657
KEYWORDS       .
SOURCE         synthetic construct
ORGANISM       other sequences; artificial sequences.
REFERENCE      1
AUTHORS        Bestetti,G., Cali,S., Orsini,G., Tonon,G., Zuffi,G. and Ghisotti,D.
TITLE          Recombinant bacterial strains for the production of natural
                  nucleosides and modified analogues thereof
JOURNAL        Patent: WO 0039307-A 5 06-JUL-2000;
                  BESTETTI GIUSEPPINA (IT) ; CALI SIMONA (IT) ; NORPHARMA SPA (IT) ;
                  ORSINI GAETANO (IT) ; TONON GIANCARLO (IT) ; ZUFFI GABRIELE (IT) ;
                  GHISOTTI DANIELA (IT)
FEATURES        Location/Qualifiers
source          1..4189
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /notes="Plasmid"
ORIGIN
Alignment Scores:
Pred. No.:      3.8e-109      Length:      4189
Score:          1222.00      Matches:      238
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 99.58%      Mismatches: 0
Query Match:    99.76%      Indels:      0
DB:             6      Gaps:      0
US-10-035-300A-2 (1-239) x AX027813 (1-4189)
Qy              1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db              1046 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTGCTGACGTAGTTTGTATGCCA 1105
Qy              21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db              1106 GCGCACCGCTGCTGCGAAGTATATTGCTGAACCTTCTCTTGAAGATGCCGGAAGTG 1165
Qy              41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db              1166 AACAAAGCTTCGCGTATGCTGGGCTTCACCGTACTTACAAAGCCGCAAAATTTCCGTA 1225
Qy              61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuLeuThrAsp 80

```

[illegible]

Search completed: July 23, 2005, 03:20:45
Job time : 5344 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 22, 2005, 22:27:52 ; Search time 677 Seconds
(without alignments)
2089.835 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225

Sequence: 1 MATPHINAEMGDFADVVLMP.....TTFNDMIKIALESVLLGDKE 239

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2.1/USPFO.spool_p/US1003300/runat_18072005_165350_27964/app_query.fasta_1.391
-DB=N_Geneseq_16Dec04 -QFWT=fascap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1003300.scgn.1.1.708 @runat_18072005_165350_27964 -NCPU=6 -ICPU=3
-NO MMAP -LARGESQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04.*

1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002as.*
7: Geneseq2002bs.*
8: Geneseq2003as.*
9: Geneseq2003bs.*
10: Geneseq2003cs.*
11: Geneseq2003ds.*
12: Geneseq2004as.*
13: Geneseq2004bs.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1225	100.0	720	8 AAD56237	Aad56237 Escherich
2	1222	99.8	720	5 AAF55440	Aaf55440 Nucleotid
3	1222	99.8	720	8 AAD54112	Aad54112 Escherich
4	1222	99.8	720	12 ADL27841	Adl27841 E coli de
5	1222	99.8	752	6 ABL51595	Ab151595 Purine nu

6	1222	99.8	3383	3 AAAS1632	Aaas1632 Plaamid p
7	1222	99.8	4189	3 AAAS1634	Aaas1634 Plaamid p
8	1222	99.8	5013	9 ACA63355	ACA63355 E. coli D
9	1222	99.8	5013	10 AAD59423	Aad59423 E. coli pu
10	1222	99.8	5013	10 ADE86127	Ade86127 E. coli p
11	1222	99.8	5495	3 AAAS1633	Aaas1633 Plaamid p
12	1222	99.8	6301	3 AAAS1635	Aaas1635 Plaamid p
13	1218	99.4	720	8 AAD56238	Aad56238 Escherich
14	1213	99.0	720	5 AAF55446	Aaf55446 Nucleotid
15	1162	94.9	732	11 ACH95836	Ach95836 Klebsiell
16	1104	90.1	3066	3 AAAS1641	Aaas1641 PGM747 co
17	1103	90.0	3165	3 AAAS1642	Aaas1642 PGM751 co
18	1103	90.0	3979	3 AAAS1643	Aaas1643 PGM800 wi
19	1103	90.0	6117	3 AAAS1644	Aaas1644 PGM807 wi
20	1101	89.9	5303	3 AAAS1636	Aaas1636 Plaamid p
21	1092	89.1	6373	3 AAAS1639	Aaas1639 pGM795 co
22	1087	88.7	6343	3 AAAS1638	Aaas1638 Plaamid p
23	1067	87.1	717	10 ACF69888	Acf69888 Photorhab
24	1067	87.1	110000	10 ACF65385_0	Acf65385 Photorhab
25	1067	87.1	110000	10 ACF67367_29	Continuation (30 o
26	1057	86.3	795	10 ADF02086	Adf02086 Bacterial
27	1017	83.0	6701	13 ADT05517	Adt05517 Haemophil
28	1017	83.0	85814	13 ADT05644	Adt05644 Haemophil
29	1013	82.7	110000	2 AAT42063_05	Continuation (6 of
30	930.5	76.0	2807	5 AAS86794	Aas86794 DNA encod
31	912	74.4	1893	5 AAS88507	Aas88507 DNA encod
32	871.5	71.1	1112	5 AAS89111	Aas89111 DNA encod
33	778	63.5	110000	6 ABA92787_5	Continuation (6 of
34	700.5	57.2	783	5 AAS86792_	Aas86792 DNA encod
35	693.5	56.6	899	13 ADR12667	Adr12667 Bacillus
36	675	55.1	843	2 AAX14178	Aax14178 H. pylori
37	671.5	54.8	110000	6 ABQ69245_19	Continuation (20 o
38	671.5	54.8	319630	6 ABQ67194_	Abq67194 Listeria
39	670.5	54.7	1171	6 ABQ70303	Abq70303 Listeria
40	670.5	54.7	110000	6 ABA03041_19	Continuation (20 o
41	659.5	53.8	2040	2 AAT77102	Aat77102 DNA encod
42	659	53.8	110000	6 ABA90521_09	Continuation (10 o
43	635	51.8	747	9 ADB09706	Adb09706 Alloiococ
44	635	51.8	110000	9 ADB12064_10	Continuation (11 o
45	607	49.6	735	10 ADC91558	Adc91558 E. faeciu

ALIGNMENTS

RESULT 1	
AAD56237	
ID	AAD56237 standard; DNA; 720 BP.
XX	
AC	AAD56237;
XX	
DT	07-AUG-2003 (first entry)
XX	
DE	Escherichia coli mutant PNP enzyme (M65V) encoding DNA #SEQ ID 1.
XX	
KW	Purine nucleoside phosphorylase; mutant; tumour; PNP; enzyme; cancer;
KW	leukaemia; glioblastoma; lymphoma; melanoma; gene therapy;
KW	gene; ds.
XX	
OS	Escherichia coli.
XX	
XX	Synthetic.
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..720
FT	/*tag= a
FT	/product= "Mutant PNP enzyme"
XX	
PN	WO2003035012-A2.
XX	
PD	01-MAY-2003.
XX	
PF	28-OCT-2002; 2002WO-US034626.
XX	
PR	26-OCT-2001; 2001US-00035300.

XX (UABR-) UAB RES FOUND.
 PA (SOUR) SOUTHERN RES INST.
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Ealick SE, Parker WB, Seerist JA, Sorscher EJ;
 XX
 DR WPI; 2003-421350/39.
 DR P-PSDB; AAE37217.
 XX
 PT Novel purified mutant purine cleaving enzyme having different biological
 PT activity than a wild-type purine cleaving enzyme, useful for treating a
 PT pathological condition characterized by abnormal cell growth.
 XX
 PS Claim 10; Page 63-64; 67pp; English.
 XX
 CC The invention relates to mutant purine nucleoside phosphorylase (PNP)
 CC enzymes and nucleic acid molecules encoding such enzymes. These mutant
 CC enzymes have greater biological activity than wild-type enzymes. PNP
 CC enzymes are useful for treating a pathological condition characterised by
 CC abnormal cell growth such as that occurring in cancer of the skin,
 CC breast, bone, colon, head or neck, kidney, larynx, lung, liver, ovary,
 CC nasopharynx, oesophagus, pancreas, prostate, bladder, rectum, stomach,
 CC thyroid, testicle and uterus as well as other conditions characterised by
 CC abnormal cell growth such as myeloid leukaemia, glioblastoma and
 CC lymphoma. Sequences of the invention are useful for treating metastatic
 CC solid tumours such as melanoma, pancreatic, liver or colonic carcinoma or
 CC to activate prodrugs in the treatment of cancer. They are useful for
 CC treating virally infected cells. They are also useful in gene therapy.
 CC The present sequence is Escherichia coli mutant PNP enzyme (M65V)
 CC encoding DNA
 XX
 SQ Sequence 720 BP; 171 A; 190 C; 187 G; 172 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.05e-139 Length: 720
 Score: 1225.00 Matches: 239
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0

US-10-035-300A-2 (1-239) x AAD56237 (1-720)

QY 1 MetAlaThrProHisIleAenAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
 DB 1 ATGGCTACCCACACATTAATGACAGAAATGGCGATTTTCGCTGACGTAGTTTGTATGCCA 60
 QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
 DB 61 GCGACCCGCTGCGTGGAGATATATGCTGAACTTTCTTGAAGATGCCGTGAAGTG 120
 QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
 DB 121 AACCAACGTTCCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCCAAAATTTCCGTA 180
 QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
 DB 181 ATGGGTCAACGGTGTGTGATCCCGTCTCTGCTCATCTACACCAAGAACATGATCACCAGAT 240
 QY 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
 DB 241 TTCGGCGTGAAGAAATATATCCGGTGGGTTCCTGTGCGCAGTTCTGCGCAGTAAATA 300
 QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
 DB 301 CTGCGCACGTCGTTATCGGTATGGTGGTGGCTGCACCGATTTCCAAAGTTAACCCGATCCGT 360
 QY 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
 DB 361 TTTAAAGACCATGACTTTGGCGGTATCGCTGACTTCGACATGGTGGCTAAGCAGATGAT 420
 QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160

DB 421 GCAGCTAAAGCACTGGGTATTGATGCTCGGTGGTAAACCTGTTCTCGCTGACCTGTC 480
 QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
 DB 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGAAAAATACGGCATTTCTCGGCGTGAA 540
 QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
 DB 541 ATGAAAGCGCTGGTATCTACGGCTGCTGCAAGATTTGGCGGAAAGCCCTGACCATC 600
 QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
 DB 601 TGCACCGTATCTGACCATCCGACTCAGAGCAGACCACTGCGCGTGGAGCGTCAGACT 660
 QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
 DB 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGTGGCGGATAAAGAG 717
 RESULT 2
 AAF55440
 ID AAF55440 standard; DNA; 720 BP.
 XX
 AC AAF55440;
 XX
 DT 29-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of a purine nucleoside phosphorylase.
 XX
 KW Deoxyribonucleoside; deoxyribose 1-phosphate; thymidine phosphorylase;
 KW purine nucleoside phosphorylase; phosphopentose mutase;
 KW phosphopentose aldolase; fructose 1,6-diphosphate aldolase;
 KW deoxyribokinase; nucleoside 2-deoxyriboseyltransferase; ss.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 CDS 1..720
 FT /tag= a
 FT /product= "purine nucleoside phosphorylase"
 XX
 WO200114566-A2.
 XX
 PD 01-MAR-2001.
 XX
 PF 18-AUG-2000; 2000WO-EP008088.
 XX
 PR 20-AUG-1999; 99EP-00116425.
 XX
 PA (HOFF) ROCHE DIAGNOSTICS GMBH.
 PA (INSP) INST PASTEUR.
 PA (PHAR-) PHARMA-WALDHOF GMBH & CO KG.
 XX
 PI Tischer W, Ihlenfeldt H, Barzu O, Sakamoto H, Pistotnik E;
 PI Marliere P, Pochet S;
 XX
 WPI; 2001-235026/24.
 P-PSDB; AAB67584.
 XX
 PT In vitro enzymatic synthesis of deoxyribonucleosides comprises reacting
 PT deoxyribose 1-phosphate and a nucleobase to form a deoxyribonucleoside
 PT and an inorganic phosphate.
 XX
 PS Disclosure; Page 44-45; 73pp; English.
 XX
 CC The present sequence encodes a purine nucleoside phosphorylase enzyme.
 CC This enzyme is involved in the biosynthesis of deoxyribonucleosides, and
 CC is used in the method of the invention. The specification describes a
 CC method for the in vitro enzymatic synthesis of deoxyribonucleosides. The
 CC method comprises reacting deoxyribose 1-phosphate and a nucleobase to
 CC form a deoxyribonucleoside and an inorganic phosphate. Enzymes which may
 CC be used in the method of the invention include thymidine phosphorylase,
 CC purine nucleoside phosphorylase, phosphopentose mutase, phosphopentose

CC aldolase, fructose 1,6-diphosphate aldolase, deoxyribokinase, and
CC nucleoside 2-deoxyribosyltransferase

SQ Sequence 720 BP; 172 A; 190 C; 187 G; 171 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,45e-139 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 5 Gaps: 0

US-10-035-300A-2 (1-239) x AAF55440 (1-720)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAATGACGAATGGCGATTCGCTGACGTAGTTTGTATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GCGCAGCCGCTCGGTGCGAAGTATATGCTGAACCTTCTTGAAGATGCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 121 AACACAGTTCCGCGGTATGCTGGGCTTCCCGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 181 ATGGGTACGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT 240
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGCGTGAAGAAATATATCGCGTGGGTTCCTGTGGCGAGTTCTCGCGCACGTAAAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGCGCGACGTCGTATCGGTATGGGTGCTGCACCGGATTCCAAAGTTAACCGCATCCGT 360
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTCCCGCTATCGCTGACATGTTGCGTACGCGATAGAT 420
Qy 141 AlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAGCACCTGGGTATGATCTCGCGTGGGTAACTGTTCTCGCGTACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACCGCGAATGTTGACGTGATGGAAAAATACGGCATTTCTCGCGGTGGA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGACGGCTGGTATCTACGGCGTCTGAGAAATTTGGCGAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCCGACTCAGCAGCAGACCACTCGCGCTGAGCGTCAGACT 660
Qy 221 ThrPheAsnAspMetIleIleValIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGCGATAAAGAG 717

RESULT 3

AAD54112

ID AAD54112 standard; DNA; 720 BP.

XX

AC AAD54112;

XX

DT 17-JUN-2003 (first entry)

XX

DE Escherichia coli cytosine nucleoside phosphorylase DNA.

XX

KW Cytosine nucleoside phosphorylase; cytosine nucleoside compound;
KW pharmaceutical; enzyme; gene; ds.
XX Escherichia coli.

XX Location/Qualifiers
FH Key 1..720
FT CDS /*tag= a
FT /product= "E. coli cytosine nucleoside phosphorylase"

XX EPI254959-A2.

XX 06-NOV-2002.

XX 01-MAY-2002; 2002EP-00253075.

XX 01-MAY-2001; 2001JP-00134352.

XX (MITA) MITSUI CHEM INC.

XX Araki T, Ikeda I, Matoishi K, Abe R, Oikawa T, Matsuba Y;

XX Ishibashi H, Nagahara K, Fukui Y;

XX WPI; 2003-158787/16.

XX P-PsDB; AAE35397.

XX Producing cytosine nucleoside compound by reacting sugar phosphate and

XX cytosine or cytosine derivative in the presence of cytosine nucleoside

XX phosphorylase.

XX Disclosure; Page 17-18; 32pp; English.

XX The invention relates to a method for producing cytosine nucleoside

XX compound by reacting sugar phosphate and cytosine or cytosine derivative

XX in the presence of cytosine nucleoside phosphorylase. The method is used

XX to produce cytosine nucleoside compounds which are intermediates for

XX pharmaceuticals. The present sequence is Escherichia coli cytosine

XX nucleoside phosphorylase DNA

XX SQ Sequence 720 BP; 172 A; 190 C; 187 G; 171 T; 0 U; 0 Other;

XX Alignment Scores:

XX Pred. No.: 2,45e-139 Length: 720

XX Score: 1222.00 Matches: 238

XX Percent Similarity: 100.00% Conservative: 1

XX Best Local Similarity: 99.58% Mismatches: 0

XX Query Match: 99.76% Indels: 0

XX DB: 8 Gaps: 0

XX US-10-035-300A-2 (1-239) x AAD54112 (1-720)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20

Db 1 ATGGCTACCCACACATTAATGACGAATGGCGATTCGCTGACGTAGTTTGTATGCCA 60

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40

Db 61 GCGCAGCCGCTCGGTGCGAAGTATATGCTGAACCTTCTTGAAGATGCCGTGAAGTG 120

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60

Db 121 AACACAGTTCCGCGGTATGCTGGGCTTCCCGGTACTTACAAAGGCCGCAAAATTTCCGTA 180

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80

Db 181 ATGGGTACGGTATGGGTATCCCGTCTGCTCCATCTACACCAAGAACTGATCACCGAT 240

Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100

Db 241 TTCGCGTGAAGAAATATATCGCGTGGGTTCCTGTGGCGAGTTCTCGCGCACGTAAAA 300

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120

XX

Db 301 CTGGCGACGCTGTTATCGGTATGGGTGCTGCACCGGATTCCTCAAGATTAAACCGCATCCGT 360
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGGCGGTATCGCTGACTTCGACATGGTGGGTAAACGCGATAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGCATGGGTATGATGCTCGGTGGGTAACTGTTCTCCGTGACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGTAATAATACGCATTCTCGCGGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGAAAGCGGTGGTATCTACGGCGTGCCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCCGCACTCAGAGCAGACCACTGCGCGTGAAGCTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 717
RESULT 4
ADL27841
ID ADL27841 standard; DNA; 720 BP.
XX
AC ADL27841;
XX
DT 20-MAY-2004 (first entry)
XX
DE E coli deoD coding sequence.
XX
KW ds; gene; selection method; yield; pathogen resistance;
KW nutritional quality; stress resistance; plant; transgenic.
XX
OS Escherichia coli.
XX
FN WO2004013333-A2.
XX
PD 12-FEB-2004.
XX
PF 18-JUL-2003; 2003WO-EP007877.
XX
PR 26-JUL-2002; 2002DE-01034287.
XX
PA (BADI) BASF PLANT SCI GMBH.
XX
PI Kock M, Frank M, Badur R;
XX
DR WPI; 2004-157134/15.
XX P-PSDB; ADL27842.
PT Method for transforming plant cells, useful for preparing transgenic
PT plants for e.g. food or production of chemicals, with selection based on
PT suppressing toxic effects of a marker protein.
XX
PS Disclosure; Page 139-140; 201pp; German.
XX
CC The present invention relates to a method for preparing transformed plant
CC cells or organisms, which comprises first transfecting a population of
CC plant cells, already containing at least one marker protein that can
CC exert a direct or indirect toxic effect, with at least one nucleic acid
CC sequence together with at least one compound able to reduce expression,
CC amount, activity and/or function of the marker protein. Transfected cells
CC that contain the nucleic acid in their genome and, because of the
CC activity of the compound have a growth advantage relative to non-
CC transformed cells are selected, under conditions where the marker protein
CC exerts its toxic effect on non-transformed cells. The method is
CC especially used to produce transgenic plants, especially those having an

CC advantageous phenotype, e.g. better nutritional quality for humans or
CC animals, production of selected chemicals or pharmaceuticals, increased
CC resistance to pathogens or environmental stress or higher yields. The
CC present sequence is a coding sequence shown in the exemplification of the
CC invention.
XX
SQ Sequence 720 BP; 172 A; 190 C; 187 G; 171 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2,45e-139 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 12 Gaps: 0
US-10-035-300A-2 (1-239) x ADL27841 (1-720)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTCGCTGACGTAGTTTTTGATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GCGACCCGCTGCGTGCAGAGTATATTGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyValGlyLysIleSerVal 60
Db 121 AACACGTTGCGCGGTATGCTGGGCTTCACCGGTACTTACAAAGCGCCCAAAATTTCCGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 181 ATGGGTCACGGTATGGGTATCCGCTCTGCTCCTACACCAAGAACTGATCCCGAT 240
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCCGGGTGAAGAAATTTATCCGCGTGGGTTCCTGTGGCGCAGTTCTTCCGCGCACGTAAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGCGCGACGTGTTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 360
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGGCGGTATCGCTGACTTCGACATGGTGGGTAAACGCGATAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGCACTGGGTATTTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAATAATACGGCATTTCTCGCGGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGAAAGCGGTGTTATCTACGCGCTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCCGCACTCAGAGCAGACCACTCCCGCTGAGCGTCAAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 717
RESULT 5
ABL51595
ID ABL51595 standard; DNA; 752 BP.
XX
AC ABL51595;
XX
DT 04-JUL-2002 (first entry)

XX Purine nucleoside phosphorylase related DNA sequence.
 XX Purine nucleoside phosphorylase; enzyme; molecular biology;
 KW biotechnology; genetic engineering; gene; ds.
 XX Escherichia coli.
 OS Synthetic.

XX Key Location/Qualifiers
 FH primer_bind 7..35
 FT /*tag= a
 FT /note= "oligonucleotide binding site"
 FT complement(709..742)
 FT /*tag= b
 FT /note= "oligonucleotide binding site"

XX RU2179188-C2.

XX 10-FEB-2002.

XX 03-MAR-2000; 2000RU-00105214.

XX 03-MAR-2000; 2000RU-00105214.

XX (ASBI=) AS RUSSIA BIO-ORGANIC CHEM INST.

XX Esipov RS, Gurevich AI, Miroshnikov AI, Chuvikovskii DV;
 PI WPI; 2002-265551/31.

XX Production of recombinant purine nucleoside phosphorylase, recombinant
 PT plasmid DNA PERPUPH01 and strain Escherichia coli BL21(DE3)perpuh01 for
 PT its realization.

XX Disclosure; Col 9-10; Opp; Russian.

XX The present invention describes a recombinant plasmid DNA PERPUPH01 which
 CC encodes an amino acid sequence of an Escherichia coli purine nucleoside
 CC phosphorylase. The recombinant plasmid DNA PERPUPH01 consists of: a
 CC NcoI/EcoRI-fragment of plasmid pET23d DNA containing a promoter and a
 CC transcription terminator of T7 RNA polymerase, a translation enhancer of
 CC gene 10 of phage T7, a beta-lactamase gene and a NcoI/EcoRI-fragment of a
 CC DNA containing the Escherichia coli purine nucleoside phosphorylase gene
 CC sequence adapted to these sites. The recombinant plasmid DNA PERPUPH01
 CC can be used in molecular biology, biotechnology and genetic engineering.
 CC The present sequence represents a nucleotide sequence given in the
 CC exemplification of the present invention

XX SQ Sequence 752 BP; 184 A; 194 C; 195 G; 179 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 2,6e-139 Length: 752
 Score: 1222.00 Matches: 238
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.58% Mismatches: 0
 Query Match: 99.76% Indels: 0
 DB: 6 Gaps: 0

US-10-035-300A-2 (1-239) x ABL51595 (1-752)
 Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
 Db 13 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGAGTAGTTTGTATGCCA 72
 Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
 Db 73 GGCGACCCGCGTGGCGAAGTATATTCGCTGAACCTTTCTTGAAGATGCCCGTGAAGTG 132
 Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
 Db 133 AACACAGTTCGCGGTATCTGGGCTTACCCGTACTTACAAAGGCCGCAAAATTTCCGTA 192

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
 Db 193 ATGGGTACCGGTATGGGTATCCGTCCTGCTCCATCTACACCAAGAACTGATCACCGAT 252
 Qy 81 PheGlyValLysTyrIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
 Db 253 TTCGCGTGAAGAAATATATCCGCGTGGTTCCTGTGGCGAGTTCCTCCCGACGTAANA 312
 Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
 Db 313 CTGCGCGAGCGTCGTTATCGGTATGGTGCCTGCACCGATTCCAAAGTTTAACCGCATCGT 372
 Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
 Db 373 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGCGTAAAGCAGTAGAT 432
 Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
 Db 433 GCAGCTAAAGCAGCTGGGTATTGATGCTCGGTGGGTAACTGTTCTCGCTGACCTGTTTC 492
 Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
 Db 493 TACTCTCGGACGCGGAAATGTTTCGACGTGATGCAAAATACGGCATTTCTCGGCGTGAA 552
 Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
 Db 553 ATGGAAGCGGCTGTTATCTACGGCTGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 612
 Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
 Db 613 TGCACCGTATCTGACCACTCCGCACTCACGAGCAGACCTGCGCGTAAAGCGTCAACT 672
 Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
 Db 673 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAG 729
 RESULT 6
 AAA51632
 ID AAA51632 standard; DNA; 3383 BP.
 AC AAA51632;
 XX 31-OCT-2000 (first entry)
 DT Plasmid pGM678 containing E. coli deod gene.
 DE Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deod;
 KW Genetically modified organism; catalyst; transglycosylation;
 KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;
 KW anti-viral; anti-tumour; mesophilic bacterium; ss.
 XX Escherichia coli.
 OS Synthetic.
 XX Key Location/Qualifiers
 FH misc_RNA 1..230
 FT /*tag= a
 FT /label= pUC18_sequence
 FT 216..952
 FT /*tag= b
 FT /product= "lacZ-deod_fused_gene"
 FT 961..3383
 FT /*tag= c
 FT /label= pUC18_sequence
 XX WO2000039307-A2.
 PN 06-JUL-2000.
 PD 23-DEC-1999; 99WO-EP010416.
 PF 23-DEC-1998; 98IT-MI002792.
 PR XX

PA (NORP-) NORPHARMA SPA.

XX Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;

XX WPI; 2000-452402/39.

XX Recombinant expression vectors used to transform cells for the production
PT of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside
PT phosphorylase.

XX Claim 14; Page 47-48; 72pp; English.

XX Host cells genetically modified to express uridine phosphorylase (UDP)
CC and purine nucleoside phosphorylase (PNP) or their corresponding crude or
CC purified extracts, either separately or in combination are used as
CC catalysts of transglycosylation reactions between a donor nucleoside and
CC an acceptor base, for preparing nucleoside analogues containing
CC heterocyclic systems with purine and/or pyrimidine bases substituted by
CC one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate
CC sugars by phosphorylase reactions and for producing nucleosides and
CC modified nucleoside analogues (all claimed). The modified or natural
CC nucleosides are used directly or as intermediates in the preparation of
CC drugs with anti-viral and anti-tumour activity and for preparing
CC oligonucleotides for therapeutic or diagnostic use. The recombinant host
CC cells are produced using recombinant plasmid expression vectors
CC comprising at least one gene sequence of a mesophilic bacterium coding
CC for a polypeptide with UDP activity and/or enzyme PNP activity and at
CC least one gene sequence coding for antibiotic resistance. In particular,
CC the E. coli udp and deoD genes are used. Recombinant strains produced
CC using the vectors express polypeptides with enzyme UDP and PNP activity
CC in large amounts, e.g. 340-1040 times higher UDP activity and 120-200
CC times higher PNP activity than non-transformed corresponding wild type
CC strains

XX SQ Sequence 3383 BP; 826 A; 868 C; 855 G; 834 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,07e-138 Length: 3383
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 3 Gaps: 0

US-10-035-300A-2 (1-239) x AAA51632 (1-3383)

Qy 1 MetAlaThrProHisIleAenAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 240 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGCTAGTTTTGATGCCA 299
Qy 21 GlyAspProLeuArgAlaIleValGlyThrPheAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 300 GGGACCCGCTGGGTGGGAAGTAGTATTGCTGGAACATTTCTTGAAGATGCCCGTGAAGTG 359
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgIleSerVal 60
Db 360 AACACGTTCCGGGTATGCTGGGCTTCACCGTACTTACAAAGCCGCAAAATTCGCTA 419
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGlyGluLeuIleThrAsp 80
Db 420 ATGGGTACCGGTATGGGTATCCGCTCCGCTGCTACACCAAGAACTGATCACCGAT 479
Qy 81 PheGlyValIleValIleIleArgValGlySerCysGlyAlaValLeuProHisValIys 100
Db 480 TTCGGCGTGAAGAAATATATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGCACGTAAAA 539
Qy 101 LeuArgAspValValIleGlyMetGlyValaCysThrAspSerIysValAsnArgIleArg 120
Db 540 CTGCGGACGCTCGTTATCGGTATGGGTGCTGACCGATTCCAAAGTTAACCGCATCCGT 599
Qy 121 PheIysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 600 TTTTAAAGACCATGACTTTGCGCTATCGCTGACTTCGACATGTTGCGTAAACGCGTAGAT 659

Qy 141 AlaAlaIysAlaLeuGlyIleAspAlaAArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 660 GCAGCTAAAGCAGCTGGGTATGTATGCTCGCGTGGGTAACCTGTTCTCCGCTACCTGTTTC 719
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyIleLeuGlyValGlu 180
Db 720 TACTCTCCGACGCGCAAAATGTTTCGACGTCGCTGCAGAAATTTGGCGCGAAGCCCTGACCATC 779
Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaIysAlaLeuThrIle 200
Db 780 ATGGAAGCGCTGATCTACGCGCTGCTGCAGAAATTTGGCGCGAAGCCCTGACCATC 839
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 840 TGCACCGTATCTGACCACTCCGCTCAGCAGACACCACTCCGCTGAGCGTACAGACT 899
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCGTTCTGCTGGCGGATAAAGAG 956
RESULT 7
AAA51634
ID AAA51634 standard; DNA; 4189 BP.
XX
AC AAA51634;
XX 31-OCT-2000 (first entry)
XX Plasmid pGM712 containing udp and deoD genes.
KW Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;
KW Genetically modified organism; catalyst; transglycosylation; anti-viral;
KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorylase;
KW anti-tumour; mesophilic bacterium; tetracycline resistance; ss.
XX Escherichia coli.
OS Synthetic.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT misc_RNA 1..242
FT /tag= a
FT /label= pUC18_sequence
FT 243..1021
FT /tag= b
FT /label= UDP_gene_sequence
FT 1022..1025
FT /tag= c
FT /label= pUC18_sequence
FT 1026..1036
FT /tag= d
FT /label= pBAD24_sequence
FT 1037..1766
FT /tag= e
FT /label= deoD_gene_sequence
FT 1767..1792
FT /tag= f
FT /label= pBAD24_sequence
FT 1793..4189
FT /tag= g
FT /label= pUC18_sequence
XX
XX WO200039307-A2.
PN
XX
XX 06-JUL-2000.
PD
XX 23-DEC-1999; 99WO-EP010416.
PF
XX 23-DEC-1998; 98IT-MI002792.
PR
XX (NORP-) NORPHARMA SPA.
PA
XX

PI Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;
 XX WPI; 2000-452402/39.
 XX Recombinant expression vectors used to transform cells for the production
 PT of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside
 PT phosphorylase.
 XX
 XX Claim 14; Page 51-52; 72pp; English.
 XX
 CC Host cells genetically modified to express uridine phosphorylase (Udp)
 CC and purine nucleoside phosphorylase (PNP) or their corresponding crude or
 CC purified extracts, either separately or in combination are used as
 CC catalysts of transglycosylation reactions between a donor nucleoside and
 CC an acceptor base, for preparing nucleoside analogues containing
 CC heterocyclic systems with purine and/or pyrimidine bases substituted by
 CC one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate
 CC sugars by phosphorolysis reactions and for producing nucleosides and
 CC modified nucleoside analogues (all claimed). The modified or natural
 CC nucleosides are used directly or as intermediates in the preparation of
 CC drugs with anti-viral and anti-tumour activity and for preparing
 CC oligonucleotides for therapeutic or diagnostic use. The recombinant host
 CC cells are produced using recombinant plasmid expression vectors
 CC comprising at least one gene sequence of a mesophilic bacterium coding
 CC for a polypeptide with UDP activity and/or enzyme PNP activity and at
 CC least one gene sequence coding for antibiotic resistance. In particular,
 CC the E. coli udp and deoB genes are used. Recombinant strains produced
 CC using the vectors express polypeptides with enzyme UDP and PNP activity
 CC in large amounts, e.g. 340-1040 times higher UDP activity and 120-200
 CC times higher PNP activity than non-transformed corresponding wild type
 CC strains
 XX
 XX Sequence 4189 BP; 998 A; 1077 C; 1081 G; 1033 T; 0 U; 0 Other;
 SQ

Alignment Scores:
 Pred. No.: 2,79e-138 Length: 4189
 Score: 1222.00 Matches: 238
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.58% Mismatches: 0
 Query Match: 99.76% Indels: 0
 DB: 3 Gaps: 0

US-10-035-300A-2 (1-239) x AAAS1634 (1-4189)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
 Db 1046 ATGCTACCCACACATTAATGCAGAAATGGCGGATTTTCGTCAGCTAGTTTTCATGCCA 1105

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
 Db 1106 GCGGACCGCTGCGTGGCAAGTATATTCCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 1165

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyValArgLysIleSerVal 60
 Db 1166 AACACAGTTCGCGGTATCTGGCTTACCGGTACTTACAAGGCCGCAAAATTTCCGTA 1225

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
 Db 1226 ATGGGTACGGTATGGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCAGT 1285

Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
 Db 1286 TTCGCGGTGAAGAAATATATCCGCTGGGTTCCTGTGGCGAGTTCTCCGCGACGTAAA 1345

Qy 101 LeuArgAspValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
 Db 1346 CTGGCGACGTCGTATCGGTATGGGTGCGTCCGACCGGATTCGAAGTTAAACCGCATCCGT 1405

Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
 Db 1406 TTTAAAGACCATGACTTTTCGCGCTATCGCTGACATGCTGCGTAAAGCGAGTAGAT 1465

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160

Db 1466 GCACCTAAAGCACCTGGGTATTGATGCTCGCGGTAACTTCTCGCGTACCTGTTTC 1525
 Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
 Db 1526 TACTCTCGGACGGCGAAATGTTCCAGCTGATGAAAAAATACGGCATTTCTGGCGTGAA 1585
 Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
 Db 1586 ATGGAAGCGGCTGGTATCTACGGGCTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 1645
 Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
 Db 1646 TGCACCTGATCTGACCACTCCGCACTCACGAGCAGACCACTGCGCGTGCAGCGTCAGACT 1705
 Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
 Db 1706 ACCTTCAACGACATGATCAAAATCGCACTGGATCCGTTCTGCTGGGCGATAAAGAG 1762

RESULT 8
 ACA63355
 ID ACA63355 standard; DNA; 5013 BP.
 XX ACA63355;
 AC ACA63355;
 XX 29-AUG-2003 (first entry)
 DT
 XX
 DE E. coli DNA encoding PNP inserted in plasmid pTRCNP.
 XX
 KW Prokaryotic host cell; transformation; transfection; vector; PNP;
 KW purine nucleoside phosphorylase; hydrolase; replicating cell;
 KW nonreplicating cell; bystander cell; purine substrate; tumour cell;
 KW virally infected cell; cancer therapy; cytostatic; mutant; pTRCNP; ds.
 XX
 OS Escherichia coli; strain JM101.
 OS Synthetic.
 XX
 FN US6491905-B1.
 XX
 PD 10-DEC-2002.
 XX
 PP 30-OCT-1998; 98US-00183188.
 XX
 PR 14-SEP-1993; 93US-00122321.
 PR 23-AUG-1996; 96US-00702181.
 PR 24-JUN-1997; 97US-00881772.
 PR 31-OCT-1997; 97US-0064676P.
 XX
 PA (UABR-) UAB RES FOUND.
 PA (SOUR) SOUTHERN RES INST.
 XX
 PI Sorscher EJ, Parker WB, Waud W, Gadi VK, Bennett LL;
 XX
 DR WPI; 2003-502635/47.
 XX
 PT Novel prokaryotic host cell for treating tumor cells and virally infected
 PT cells transformed or transfected by a vector comprising a DNA sequence
 PT encoding Escherichia coli derived purine nucleoside phosphorylase.
 XX
 PS Claim 1; Fig 15; 45pp; English.
 XX
 CC The present invention relates to a prokaryotic host cell transformed or
 CC transfected by a vector comprising a DNA sequence encoding Escherichia
 CC coli derived purine nucleoside phosphorylase (PNP) or hydrolase. The
 CC vector is useful for killing targeted replicating or nonreplicating
 CC mammalian cells, and bystander cells. The host cell is useful in
 CC combination with a purine substrate for treating tumour and/or virally
 CC infected cells. The vector is useful in cancer therapy. The present
 CC sequence represents DNA encoding E.coli PNP inserted in plasmid pTRCNP
 SQ Sequence 5013 BP; 1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;

Alignment Scores:

```
Pred. No.: 3.57e-138 Length: 5013
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 9 Gaps: 0

US-10-035-300A-2 (1-239) x ACA63355 (1-5013)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 413 ATGGCTACCCACACATTAATGCAGAAATGGCGCATTCGCTGACGTAGTTTTGATGCCA 472

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 473 GCGACCGCTGCGTGGAGTATATTCGTGAACCTTCCTTGAAGATGCCCGTGAAGTG 532

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 533 AACAAAGTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 592

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 593 ATGGGTACCGGTATGGGTATCCCGTCTCTCCATCTACACCAAGAACTGATCACCAGT 652

Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 653 TTCGCGTGAAGAAATATTCGCGTGGTTCCTGTGGCGGAGTTCCTGCGCAGTAAAA 712

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 713 CTGCGCAGCTGCTATCGGTATGGGTGCTGCGACCGATTCCAAAGTTAACCGCATCCGT 772

Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 773 TTTAAAGACCATGACTTTGCGCGCTATCGCTGACTTCGACATGGTGCCTAAACGCGATGAT 832

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 833 GCAGCTAAAGCATGGGTATGTATGCTCGGTGGGTAACTGTCTTCGCTGACCTGTTTC 892

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 893 TACTCTCCGACGGCGAAATGTTTCGACGTGATGGAATAATACGCCATTCCTCGGCGTGA 952

Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 953 ATGGAAGCGCTGCTATCTACGCGCTCGCTGCAGAAATTTGCGCGGAAGCCCTGACCATC 1012

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1013 TGCACCGTATCTGACCACTCCGACCTCAGACGACGACCACTGCGCGTGAAGCTCAGACT 1072

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 1129

RESULT 9
AAD59423
ID AAD59423 standard; DNA; 5013 BP.
XX
AC
AD59423;
XX
XX
18-DEC-2003 (first entry)
XX
DE E.coli purine analogue nucleoside phosphorylase DNA.
XX
KW Cytostatic; virucide; purine cleavage enzyme; tumour; therapy; PNP;
KW drug screening; purine analogue nucleoside phosphorylase; ds.
XX
OS Escherichia coli.
XX
PN US2003077268-A1.
XX
```

```
XX 24-APR-2003.
PD
XX
XX 18-JUL-2002; 2002US-00198034.
PF
XX
XX 14-SEP-1993; 93US-00122321.
PR
XX 23-AUG-1996; 96US-00702181.
PR
XX 24-JUN-1997; 97US-00881772.
PR
XX 31-OCT-1997; 97US-0064676P.
PR
XX 30-OCT-1998; 98US-00183188.
XX
XX (SORS/) SORSCHER E J.
PA (PARK/) PARKER W B.
PA (WAUD/) WAUD W.
PA (GADI/) GADI V K.
PA (BENN/) BENNETT L L.
XX
XX Sorscher EJ, Parker WB, Waud W, Gadi VK, Bennett LL;
XX WPI; 2003-755064/71.
XX
XX Killing (non-)replicating targeted mammalian cells and bystander cells
XX useful for treating tumor cells and/or virally infected cells comprises
XX delivering purine nucleotide phosphorylase or hydrolase cleavage enzyme
XX to cells.
XX
XX Disclosure; Page 20-23; 40pp; English.
XX
XX The present invention relates to a method of killing replicating or non-
XX replicating, targeted mammalian cells and bystander cells, comprising
XX delivering a purine cleavage enzyme to the targeted mammalian cells and
XX contacting the targeted cells with a purine cleavage enzyme substrate to
XX kill the targeted cells when cleaved by the enzyme. The purine cleavage
XX enzyme is used for killing replicating or non-replicating targeted
XX mammalian cells and bystander cells. They are used for treating tumour
XX cells and virally infected cells. The method may also be useful in drug
XX screening. This sequence represents Escherichia coli PNP (purine analogue
XX nucleoside phosphorylase) DNA
XX
XX Sequence 5013 BP; 1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 3.57e-138 Length: 5013
XX Score: 1222.00 Matches: 238
XX Percent Similarity: 100.00% Conservative: 1
XX Best Local Similarity: 99.58% Mismatches: 0
XX Query Match: 99.76% Indels: 0
XX DB: 10 Gaps: 0
XX
XX US-10-035-300A-2 (1-239) x AAD59423 (1-5013)
XX
XX Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
XX Db 413 ATGGCTACCCACACATTAATGCAGAAATGGCGCATTCGCTGACGTAGTTTTGATGCCA 472
XX
XX Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
XX Db 473 GCGACCGCTGCGTGGAGTATATTCGTGAACCTTCCTTGAAGATGCCCGTGAAGTG 532
XX
XX Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
XX Db 533 AACAAAGTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 592
XX
XX Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
XX Db 593 ATGGGTACCGGTATGGGTATCCCGTCTCTCCATCTACACCAAGAACTGATCACCAGT 652
XX
XX Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
XX Db 653 TTCGCGTGAAGAAATATTCGCGTGGTTCCTGTGGCGGAGTTCCTGCGCAGTAAAA 712
XX
XX Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
XX Db 713 CTGCGCAGCTGCTATCGGTATGGGTGCTGCGACCGATTCCAAAGTTAACCGCATCCGT 772
XX
XX Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
XX Db 773 TTTAAAGACCATGACTTTGCGCGCTATCGCTGACTTCGACATGGTGCCTAAACGCGATGAT 832
XX
XX Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
XX Db 833 GCAGCTAAAGCATGGGTATGTATGCTCGGTGGGTAACTGTCTTCGCTGACCTGTTTC 892
XX
XX Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
XX Db 893 TACTCTCCGACGGCGAAATGTTTCGACGTGATGGAATAATACGCCATTCCTCGGCGTGA 952
XX
XX Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
XX Db 953 ATGGAAGCGCTGCTATCTACGCGCTCGCTGCAGAAATTTGCGCGGAAGCCCTGACCATC 1012
XX
XX Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
XX Db 1013 TGCACCGTATCTGACCACTCCGACCTCAGACGACGACCACTGCGCGTGAAGCTCAGACT 1072
XX
XX Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
XX Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 1129
XX
XX RESULT 9
XX AAD59423
XX ID AAD59423 standard; DNA; 5013 BP.
XX XX
XX AC
XX AD59423;
XX XX
XX XX
XX 18-DEC-2003 (first entry)
XX XX
XX DE E.coli purine analogue nucleoside phosphorylase DNA.
XX XX
XX KW Cytostatic; virucide; purine cleavage enzyme; tumour; therapy; PNP;
XX KW drug screening; purine analogue nucleoside phosphorylase; ds.
XX XX
XX OS Escherichia coli.
XX XX
XX PN US2003077268-A1.
XX
```

Db	713	CTGGCGAGCTGCTGTTATCGGTATCGGTGCTGCACCGATTCCAAAGTTAAACCGCATCCGT	777
Qy	121	PhelysAspHisAspPheAlaAlaileAlaAspPheAspMetValArgAsnAlaValAsp	140
Db	773	TTTAAAGACCATGACTTTGCCGCTATCGCTGACTTCGACATGGTGGCTAACGCAGTAGAT	832
Qy	141	AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe	160
Db	833	GCACCTAAAGCACATGGGTATTGATGCTCGCTGGGTAACTTGTTCTCGCTGACCTGTTCT	892
Qy	161	TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu	180
Db	893	TACTCTCCGAGCGCGAAATGTTGCAGCTGATGAGAAATAACGGCATTTCTCGGCGTGAA	952
Qy	181	MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle	200
Db	953	ATGGAAGCGCTGGTATCTACGGCGTGCCTGCAGAAATTGGCGCGNAAGCCCTGCACCATC	1012
Qy	201	CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr	220
Db	1013	TGCACCGCTATCTGACCACATCCGCACTCAGCAGCAGACCACTGCCTGAGCGTCAGACT	1072
Qy	221	ThrPheAspMetIleLeuIleAlaLeuGluSerValLeuLeuGlyAspLysGlu	239
Db	1073	ACCTTCAACGACATGATCAAAATCGACTGGATCCGTTCTGCTGGGCGATAAAGAG	1129
RESULT 10			
AD	AD86127	ID	AD86127 standard; DNA; 5013 BP.
XX	AC	AD86127;	
XX	XX	29-JAN-2004 (first entry)	
XX	DE	E. coli purine nucleoside phosphorylase (PNP) gene.	
KW	KW	Tumour growth inhibition; purine analogue; stable transformation;	
KW	KW	prokaryotic host cell; purine nucleoside phosphorylase; PNP;	
KW	KW	purine nucleoside hydrolase; tumour cell; virally infected cell;	
KW	KW	toxin gene therapy; cytostatic; virucide; gene; ds.	
OS	OS	Escherichia coli.	
XX	XX	US2003134819-A1.	
XX	PD	17-JUL-2003.	
XX	PF	18-JUL-2002; 2002US-00198228.	
XX	PR	14-SEP-1993; 93US-00122321.	
XX	PR	23-AUG-1996; 96US-00702181.	
XX	PR	24-JUN-1997; 97US-00881772.	
XX	PR	31-OCT-1997; 97US-0084676P.	
XX	XX	30-OCT-1998; 98US-00183188.	
PA	PA	(SORS/) SORSCHER E J.	
PA	PA	(PARK/) PARKER W B.	
PA	PA	(WAUD/) WAUD W.	
PA	PA	(GADI/) GADI V K.	
PA	PA	(BENN/) BENNETT L L.	
XX	PI	Sorscher EJ, Parker WB, Waud W, Gadi VK, Bennett LL;	
XX	DR	WPI; 2003-897265/82.	
XX	PT	Use of purine analogs for inhibiting tumor growth and treating virally	
XX	PT	infected cells.	
PS	PS	Example 25; Page 20-23; 40pp; English.	
CC	CC	The present invention relates to a method of inhibiting tumour growth by	
CC	CC	administration of a purine analogue. The method comprises the stable	
CC	CC	transformation of a prokaryotic host cell with a vector including a DNA	

CC	sequence for purine nucleoside phosphorylase (PNP) or hydrolase. The
CC	transformed prokaryotic host cell is used in combination with a purine
CC	substrate to treat tumour and/or virally infected cells. The method
CC	provides toxin gene therapy to overcome the problem of inefficient
CC	delivery, cell replication dependent killing and low toxin diffusion
CC	between cells. The present sequence represents the Escherichia coli
CC	purine nucleoside phosphorylase (PNP) gene.
XX	
SQ	Sequence 5013 BP; 1231 A; 1258 C; 1325 G; 1199 T; 0 U; 0 Other;
 Alignment Scores:	
Pred. No.:	3.57e-138 Length: 5013
Score:	1222.00 Matches: 238
Percent Similarity:	100.00% Conservative: 1
Best Local Similarity:	99.58% Mismatches: 0
Query Match:	99.76% Indels: 0
DB:	10 Gaps: 0
 US-10-035-300A-2 (1-239) x ADE86127 (1-5013)	
Qy	1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaaspValLeuMetPro 20
Db	413 ATGGCTACCCACCACATTAATGCAGAAATGGCGATTTCGTGACGTAGTTTTCGATGCCA 472
Qy	21 GlyAspProLeuArgAlaLysTyrrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db	473 GGCGACCGCTGGCTGCGAAGTAGTATTGCTGAACATTTTCCTTGAGAAGATGCCCGTGAAGTG 532
Qy	41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrrysGlyArgLyelleSerVal 60
Db	533 AACAAACGTTTCGGCGATGCTGGGGCTTCACCGGTACTTACAAGGCCGCAAAATTTCCGTA 592
Qy	61 MetGlyHisGlyValGlyIleProSerCysSerileTyrrThryvsGluLeulleThrAsp 80
Db	593 ATGGGTACCGGTATGGGTATCCCGTCTCGTCCATCTACACCAAGAAGTATGATCACCGAT 652
Qy	81 PheGlyValLyslysIlelleArvgalGlyserCysGlyAlaValLeuProHisValLys 100
Db	653 TTGGCGGTGAAGAAAATAATATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCCGACGTAAAA 712
Qy	101 LeuArgAspValValllleGlyMetGlyAlaCysThrAspSerLysValAsnAtqIleArg 120
Db	713 CTGGCGACGTCGTATTCGGGTATGGGTGCTGCACCGCATTCCTCAAAGTTAACCCGATCCGT 772
Qy	121 PhelysAspHisaspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db	773 TTTTAAAGACCATGACTTTGGCGGTATCGCTGACTTTCGACATGGTGGCTAACGCAAGTAGAT 832
Qy	141 AlaAlaylsAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db	833 GCAGCTAAAGCACTGGGTATTGATGCTCGCGTGGGTAACTGTTTCTCCGCTGACCTGTTTC 892
Qy	161 TyrSerProAspGlyGluMetPheAspValMetGlubLysTyrrGlyIleLeuGlyValGlu 180
Db	893 TACTCTCCGACGCGCAAAATGTTTCGACGTGATGGAATAATACGCCATTTCTCGCGGTGGAA 952
Qy	181 MetGluAlaalaGlyIleTyrrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db	953 ATGGAACGCGCTGGTATCTACGGGCTCGCTGCAGAAATTTGGCGCAAAAGCCCTGACCATC 1012
Qy	201 CysThrValSerAspHisIleArgrHrisGluGlnThrThrAlaalaGluArgGlnThr 220
Db	1013 TGCAACGCTATCTGACCAATCCGACATCCGACATCACGAGCAGACCACTGCCGCTGAGCGTCAGACT 1072
Qy	221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db	1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATTCGCTTTCGTGGGCGGATAAAGAG 1129
 RESULT 11	
AAA51633	
ID	AAA51633 standard; DNA; 5495 BP.
XX	
AC	AAA51633;

```

XX 31-OCT-2000 (first entry)
XX Plasmid pQM707 containing deoD and tetracycline resistance genes.
XX
XX Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoB;
XX Genetically modified organism; catalyst; transglycosylation; anti-viral;
XX nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;
XX anti-tumour; mesophilic bacterium; tetracycline resistance; ss.
XX
XX Escherichia coli.
XX Synthetic.
XX Chimeric.
XX
XX Key Location/Qualifiers
XX misc_RNA 1..230
XX /tag= a
XX /label= pUC18_sequence
XX CDS 216..952
XX /tag= b
XX /product= "lacZ-udp_fused_gene"
XX misc_RNA 961..978
XX /tag= c
XX /label= pUC18_sequence
XX misc_RNA 979..1422
XX /tag= d
XX /label= pHP45_sequence
XX misc_RNA 1423..2822
XX /tag= e
XX /label= pBR322_Tet_gene_sequence
XX misc_RNA 2823..3090
XX /tag= f
XX /label= pHP45_sequence
XX misc_RNA 3091..5455
XX /tag= g
XX /label= pUC18_sequence
XX
XX WO200039307-A2.
XX
XX 06-JUN-2000.
XX
XX 23-DEC-1999; 99WO-EP010416.
XX
XX 23-DEC-1998; 98IT-MI002792.
XX (NORP-) NORPHARMA SPA.
XX
XX Bestetti G, Cali' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;
XX WPI; 2000-452402/39.
XX
XX Recombinant expression vectors used to transform cells for the production
XX of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside
XX phosphorylase.
XX
XX Claim 14; Page 48-50; 72pp; English.
XX
XX Host cells genetically modified to express uridine phosphorylase (Udp)
XX and purine nucleoside phosphorylase (PNP) or their corresponding crude or
XX purified extracts, either separately or in combination are used as
XX catalysts of transglycosylation reactions between a donor nucleoside and
XX an acceptor base, for preparing nucleoside analogues containing
XX heterocyclic systems with purine and/or pyrimidine bases substituted by
XX one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate
XX sugars by phosphorolysis reactions and for producing nucleosides and
XX modified nucleoside analogues (all claimed). The modified or natural
XX nucleosides are used directly or as intermediates in the preparation of
XX drugs with anti-viral and anti-tumour activity and for preparing
XX oligonucleotides for therapeutic or diagnostic use. The recombinant host
XX cells are produced using recombinant plasmid expression vectors
XX comprising at least one gene sequence of a mesophilic bacterium coding
XX for a polypeptide with UDP activity and/or enzyme PNP activity and at
XX least one gene sequence coding for antibiotic resistance. In particular,

```

```

CC the E. coli udp and deoD genes are used. Recombinant strains produced
CC using the vectors express polypeptides with enzyme udp and PNP activity
CC in large amounts, e.g. 340-1040 times higher UDP activity and 120-200
CC times higher PNP activity than non-transformed corresponding wild type
CC strains
XX
XX Sequence 5495 BP; 1335 A; 1416 C; 1460 G; 1284 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 4.05e-138 Length: 5495
XX Score: 1222.00 Matches: 238
XX Percent Similarity: 100.00% Conservative: 1
XX Best Local Similarity: 99.58% Mismatches: 0
XX Query Match: 99.76% Indels: 0
XX DB: 3 Gaps: 0
XX
XX US-10-035-300A-2 (1-239) x AA51633 (1-5495)
XX
XX QY 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
XX DB 240 ATGGCTACCCACACATTATATGAGAAATGGCGATTTCCTGACGTAGTTTGTATGCCA 299
XX
XX QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
XX DB 300 GGGCAGCCGCTGGGTGCGAAGTATATTGCTGAAACCTTTCTTGAAGATGCCGTGAAGTG 359
XX
XX QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
XX DB 360 AACACGTTCCGGGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCCAAATTTCCGTA 419
XX
XX QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
XX DB 420 ATGGGTACCGGTATGGGTATCCGCTCTCTCCATCTACACCAAGAACTGATCACCAGT 479
XX
XX QY 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
XX DB 480 TTCGGCGTGAAGAAATATATCCGCTGGGTTCCTGTGGCGCAGTTCCTGGCGCACGTAAAA 539
XX
XX QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
XX DB 540 CTGGCGACGTCGTATTCGGTATGGGTGCTGCACCGATTCCAAAGTTAACCGCATCCGT 599
XX
XX QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
XX DB 600 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGCTGCGTAACGCAGTAGAT 659
XX
XX QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
XX DB 660 GCAGCTAAAGCACTGGGTATTTGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCTGTT 719
XX
XX QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
XX DB 720 TATCTCCGACCGCGGAAATGTTGACGCTGATGGAAAAAATACGCATTCCTCGCGGTGAA 779
XX
XX QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
XX DB 780 ATGGAAGCGCTGGTATCTACGGCGTGGTGCAGAAATTTGGCGGAAAGCCCTGACCATC 839
XX
XX QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
XX DB 840 TGCACCGGTATCTGACCACTCCGCACTCAGCAGCAGACCACTCCGCTGAGCGTCAGACT 899
XX
XX QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
XX DB 900 ACCTTCACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGGCGGATAAAGAG 956
XX
XX RESULT 12
XX AA51635
XX ID AA51635 standard; DNA; 6301 BP.
XX
XX AC AA51635;
XX
XX 31-OCT-2000 (first entry)

```


XX DE Plasmid pGM716 containing udp, deoD and tetracycline resistance genes.

XX KW Uridine phosphorylase; udp; purine nucleoside phosphorylase; PNP; deoD;

XX KW Genetically modified organism; catalyst; transglycosylation; anti-viral;

XX KW nucleoside analogue; alpha-pentose-1-phosphate sugar; phosphorolysis;

XX KW anti-tumour; mesophilic bacterium; tetracycline resistance; ss.

XX OS Escherichia coli.

XX OS Synthetic.

XX OS Chimeric.

XX FH Key Location/Qualifiers

XX FT misc_RNA 243..1021

XX FT /tag= a

XX FT /label= udp_gene

XX FT misc_RNA 1037..1766

XX FT /tag= b

XX FT /label= deoD_gene

XX FT misc_RNA 2229..3628

XX FT /tag= c

XX FT /label= pBR322_Tet_gene_sequence

XX PN WO200039307-A2.

XX PD 06-JUL-2000.

XX PP 23-DEC-1999; 99WO-EP010416.

XX PR 23-DEC-1998; 98IT-MI002792.

XX PA (NORP-) NORPHARMA SPA.

XX PI Bestetti G, Calli' S, Ghisotti D, Orsini G, Tonon G, Zuffi G;

XX DR WPI; 2000-452402/39.

XX PT Recombinant expression vectors used to transform cells for the production

XX PT of e.g. nucleosides encode uridine phosphorylase and/or purine nucleoside

XX PT phosphorylase.

XX PS Claim 14; Page 53-55; 72pp; English.

XX CC Host cells genetically modified to express uridine phosphorylase (Udp)

XX CC and purine nucleoside phosphorylase (PNP) or their corresponding crude or

XX CC purified extracts, either separately or in combination are used as

XX CC catalysts of transglycosylation reactions between a donor nucleoside and

XX CC an acceptor base, for preparing nucleoside analogues containing

XX CC heterocyclic systems with purine and/or pyrimidine bases substituted by

XX CC one or more nitrogen atoms, for preparing alpha-pentose-1-phosphate

XX CC sugars by phosphorolysis reactions and for producing nucleosides and

XX CC modified nucleoside analogues (all claimed). The modified or natural

XX CC nucleosides are used directly or as intermediates in the preparation of

XX CC drugs with anti-viral and anti-tumour activity and for preparing

XX CC oligonucleotides for therapeutic or diagnostic use. The recombinant host

XX CC cells are produced using recombinant plasmid expression vectors

XX CC comprising at least one gene sequence of a mesophilic bacterium coding

XX CC for a polypeptide with udp activity and/or enzyme PNP activity and at

XX CC least one gene sequence coding for antibiotic resistance. In particular,

XX CC the E. coli udp and deoD genes are used. Recombinant strains produced

XX CC using the vectors express polypeptides with enzyme Udp and PNP activity

XX CC in large amounts, e.g. 340-1040 times higher Udp activity and 120-200

XX CC times higher PNP activity than non-transformed corresponding wild type

XX CC strains

XX SQ Sequence 6301 BP; 1507 A; 1625 C; 1686 G; 1483 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4,89e-138 Length: 6301

Score: 1222.00 Matches: 238

Percent Similarity: 100.00% Conservative: 1

Best Local Similarity: 99.58% Mismatches: 0

Query Match: 99.76% Indels: 0

DB: 3 Gaps: 0

US-10-035-300A-2 (1-239) x AAA51635 (1-6301)

QY 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20

DB 1046 ATGGCTACCCACACATTAATGCAAGAAATGGCGATTTCGCTGACGTAGTTTGATGCCA 1105

QY 21 GlyAspProLeuArgAlaIleTyzIleAlaGluThrPheLeuGluAspAlaArgGluVal 40

DB 1106 GCGACCCCGCTGCGTGGCAAGTATATTCCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 1165

QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgIlysIleSerVal 60

DB 1166 AACAAACGTTCCGGGTATGCTGGGCTTCACCGGTACTTACAAAGGCGCAAAATTTCCGTA 1225

QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsThrLysGluLeuIleThrAsp 80

DB 1226 ATGGGTACCGGTATGGGTATCCCGTCTCTGCTCATCTACCAAGAACTGATCACCGAT 1285

QY 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100

DB 1286 TTCGGGTGAAGAAATATTCGGGTGGTTCCTGTGGCGAGTTCTGCCGACACGTAAA 1345

QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120

DB 1346 CTGGCGGACGTCGTTATCGGTATGGTGCCTGCACCGATTCCAAAGTTAACCGCATCGGT 1405

QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140

DB 1406 TTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGCGTAACGACATAGAT 1465

QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160

DB 1466 GCAGCTAAAGCACTGGGTATTGATGCTGCGTGGGTAACTGTTCTCCGCTGACCTGTC 1525

QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyIleLeuGlyValGlu 180

DB 1526 TACTCTCCGGACGGCGAAATGTTCCACGTGATGGAATAATACGGCATTTCTCGGCTGGAA 1585

QY 181 MetGluAlaAlaGlyIleTyrlsValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200

DB 1586 ATGGAAGCGGTGGTATCTACGGCGTCTGCGAATTTGGCGGAAGCCCTGACCATC 1645

QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220

DB 1646 TGCACCGTATTGACCATCTCCGACATCCGACGACAGACCATGCGCTGAGCGTCAGACT 1705

QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239

DB 1706 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAG 1762

RESULT 13

AAD56238

ID AAD56238 standard; DNA; 720 BP.

XX AAD56238;

XX AAD56238;

DT 07-AUG-2003 (first entry)

XX Escherichia coli mutant PNP enzyme (A157V) encoding DNA #SEQ ID 3.

DE Purine nucleoside phosphorylase; mutant; tumour; PNP; enzyme; cancer;

XX leukaemia; glioblastoma; lymphoma; melanoma; carcinoma; gene therapy;

KW gene; ds.

KW Escherichia coli.

OS Synthetic.

XX Location/Qualifiers

XX Key 1..720

FT CDS

FT /tag= a

FT /product= "Mutant PNP enzyme"

CC The present sequence encodes a purine nucleoside phosphorylase enzyme.
CC This enzyme is involved in the biosynthesis of deoxyribonucleosides, and
CC is used in the invention. The specification describes a
CC method for the in vitro enzymatic synthesis of deoxyribonucleosides. The
CC method comprises reacting deoxyribose 1-phosphate and a nucleobase to
CC form a deoxyribonucleoside and an inorganic phosphate. Enzymes which may
CC be used in the method of the invention include thymidine phosphorylase,
CC purine nucleoside phosphorylase, phosphopentose mutase, phosphopentose
CC aldolase, fructose 1,6-diphosphate aldolase, deoxyribokinase, and
CC nucleoside 2-deoxyribosyltransferase
XX

SQ Sequence 720 BP; 174 A; 190 C; 185 G; 171 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.08e-138 Length: 720
Score: 1213.00 Matches: 236
Percent Similarity: 99.58% Conservativity: 2
Best Local Similarity: 98.74% Mismatches: 1
Query Match: 99.02% Indels: 0
DB: 5 Gaps: 0

US-10-035-300A-2 (1-239) x AAF55446 (1-720)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAAATGCGAATGGCGATTTTCGCTGACGTAGTTTGTATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GCGACCCGCTGCGTGGCAAGTATATCTGTAACCTTCTTGAAGATGCCCGTGAAGTG 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 121 AACACGTTCCGCGTATGCTGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuThrAsp 80
Db 181 ATGGGTACGGTATGGGTATCCCGTCCGCTCCATCTACCAAGAACTGATCACCAGT 240
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGGCGTGAAGAAATATATCCGGTGGGTTCCTGTGGCGAGTTCTGCCGACGTAATA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGCGCGCAGTCGTATATCGGTATGGGTACCTGACCGCATTCCAAAGTTTAAACCGCATCGT 360
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAGACCAATGACTTTGGCGCTATCGCTGACTTGGACATGGTGGCGTAACGCGATGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAGACACTGGGTATGTATGCTGCGTGGGTAACTTCTCGCTGACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCCGACGGCGAAATGTTCCAGCTGATGGAATAATACGGCATTTCTCGCGGTGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGAAAGCGGTGATATCTACGGCTCGCTCGAATTTGCGCGAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCATCTCCGACTCACGAGCAGACCATCTGCGGTGACGCTGACACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATATAAGAG 717

RESULT 15
ACH95836

ID ACH95836 standard; DNA; 732 BP.

XX ACH95836;

XX 29-JUL-2004 (first entry)

XX Klebsiella pneumoniae polynucleotide seqid 1631.

XX Recombinant expression vector; transcription regulatory element;

KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.

XX Klebsiella pneumoniae.

XX US6610836-B1.

XX 26-AUG-2003.

XX 27-JAN-2000; 2000US-00489039.

XX 29-JAN-1999; 99US-0117747P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL, Osborne M;

XX WPI; 2003-895346/82.

XX P-ESDB; ABO62285.

XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for
preparing a vaccine composition against Klebsiella pneumoniae.

XX Disclosure; SEQ ID NO 1631; 932pp; English.

XX The invention describes a new isolated nucleic acid encoding a Klebsiella
pneumoniae polypeptide. Also described are: a recombinant expression
vector comprising the nucleic acid, operably linked to a transcription
regulatory element; and a cell comprising the recombinant expression
vector. The nucleic acid is useful for preparing a vaccine composition
against Klebsiella pneumoniae. This sequence encodes a Klebsiella
pneumoniae polypeptide of the invention

SQ Sequence 732 BP; 167 A; 200 C; 221 G; 144 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 5.33e-132 Length: 732
Score: 1162.00 Matches: 226
Percent Similarity: 97.07% Conservativity: 6
Best Local Similarity: 94.56% Mismatches: 7
Query Match: 94.86% Indels: 0
DB: 11 Gaps: 0

US-10-035-300A-2 (1-239) x ACH95836 (1-732)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 13 ATGCACTCTCTCACTTAACGCAAGAAATGGCGATTTTCGCTGACGTGTTTATGCCG 72
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 73 GCGACCCGCTGCGCGGAGACATTTGCAGAACTTCTCGAAGACGTCGCTGAAGTG 132
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 133 AACACGTCGCGCATGCTGGGCTTCCCGGTACCTATTAAGCCGCAAAATCTCCGTG 192
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuThrAsp 80
Db 193 ATGGGCGACGGATGGGGATCCCGTCTCTATCTACACCAAGAGCTGATCACCAGT 252
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 253 TTCGGCGTGAAGAAATATATCCCGTGGGTCTTCTGCGCGCGGTGCGTGAAGATGTCANA 312

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 23, 2005, 01:24:48 ; Search time 220 Seconds
(without alignments)
1777.592 Million cell updates/sec

Title: US-10-035-300A-2
Perfect score: 1225
Sequence: 1 MATPHINAEMGDFADVVLMP.....TTFNMIKIALESVLLGDKE 239

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10035300/runat_18072005_165351_28002/app.query.fasta_1.391
-DB=Issued_Patents_NA -QPM=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -UNITS=bits -START=1 -END=1 -MATRIX=blomsm2 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=ptc -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-USER=US10035300 @CGN 1 1 105 @runat_18072005_165351_28002 -NCPU=6 -ICPU=3
-NO MAP -LARGEROVER -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1222	99.8	5013	US-09-183-188B-5	Sequence 5, Appli
2	1162	94.9	732	US-09-489-039A-1631	Sequence 1631, Ap
3	1057	86.3	795	US-09-543-681A-2371	Sequence 2371, Ap
4	1013	82.7	1830121	US-09-557-884-1	Sequence 1, Appli
5	1013	82.7	1830121	US-09-643-990A-1	Sequence 1, Appli
6	778	63.5	640681	US-09-790-988-1	Sequence 1, Appli
7	728	59.4	744	US-09-902-540-3568	Sequence 3568, Ap
8	728	59.4	19019	US-09-902-540-1171	Sequence 1171, Ap
9	607	49.6	735	US-09-107-532A-1185	Sequence 1185, Ap
10	591	48.2	714	US-09-134-000C-2608	Sequence 2608, Ap
11	576.5	47.1	711	US-09-583-110-37	Sequence 37, Appl
12	576.5	47.1	714	US-09-107-433-1967	Sequence 1967, Ap

13	575.5	47.0	13440	3	US-08-961-527-128	Sequence 128, App
14	572	46.7	732	3	US-09-134-001C-1402	Sequence 1402, Ap
15	569	46.4	2962	4	US-09-710-279-3511	Sequence 3511, Ap
16	569	46.4	5301	4	US-08-956-171B-443	Sequence 443, App
17	569	46.4	5301	4	US-08-781-986A-443	Sequence 443, App
18	557.5	45.5	2006	4	US-08-956-171B-378	Sequence 378, App
19	557.5	45.5	2006	4	US-08-781-986A-378	Sequence 378, App
20	538.5	44.0	5739	4	US-09-634-238-1	Sequence 1, Appli
21	466	38.0	580073	4	US-08-545-528D-1	Sequence 1, Appli
22	300	24.5	1124	4	US-08-809-254A-3	Sequence 3, Appli
23	245.5	20.0	897	4	US-09-489-039A-860	Sequence 860, App
24	243.5	19.9	22108	3	US-09-053-197A-3	Sequence 3, Appli
25	243.5	19.9	22108	3	US-09-085-761A-3	Sequence 3, Appli
26	231	18.9	1830121	4	US-09-557-884-1	Sequence 1, Appli
27	231	18.9	1830121	4	US-09-643-990A-1	Sequence 1, Appli
28	227.5	18.6	798	4	US-09-543-681A-1537	Sequence 1537, Ap
29	167	13.6	3072	3	US-09-221-017B-337	Sequence 337, App
30	139	11.3	711	4	US-09-107-532A-3162	Sequence 3162, Ap
31	136	11.1	717	4	US-09-489-039A-5257	Sequence 5257, Ap
32	131	10.7	729	4	US-09-543-681A-1894	Sequence 1894, Ap
33	130	10.6	699	3	US-09-347-878-17	Sequence 17, Appl
34	130	10.6	699	4	US-09-711-164-226	Sequence 226, App
35	122.5	10.0	711	4	US-09-134-000C-1699	Sequence 1699, Ap
36	116.5	9.5	927	4	US-09-328-352-534	Sequence 534, App
37	115	9.4	687	4	US-09-710-279-2935	Sequence 2935, Ap
38	115	9.4	687	4	US-09-710-279-3195	Sequence 3195, Ap
39	115	9.4	702	3	US-09-134-001C-806	Sequence 806, App
40	115	9.4	3781	4	US-09-710-279-3993	Sequence 3993, Ap
41	115	9.4	3805	4	US-09-710-279-3687	Sequence 3687, Ap
42	110	9.0	4104	3	US-08-961-527-169	Sequence 169, App
43	108	8.8	693	4	US-09-583-110-2551	Sequence 2551, Ap
44	108	8.8	720	4	US-09-107-433-826	Sequence 826, App
45	106	8.7	765	4	US-09-583-110-2358	Sequence 2358, Ap

ALIGNMENTS

RESULT 1

US-09-183-188B-5
; Sequence 5, Application US/09183188B
; Patent No. 6491905
; GENERAL INFORMATION:
; APPLICANT: Sorscher, Eric J.
; APPLICANT: Parker, William B.
; APPLICANT: Gadi, Vijayakrishna K.
; TITLE OF INVENTION: RECOMBINANT E. COLI FOR DELIVERY OF PNP TO TUMOR CELLS
; FILE REFERENCE: UAB-12405/22
; CURRENT APPLICATION NUMBER: US/09/183,188B
; PRIOR FILING DATE: 1998-10-30
; PRIOR FILING DATE: 1996-08-23
; PRIOR FILING DATE: 1993-09-14
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 5013
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-183-188B-5
Alignment Scores:
Pred. No.: 5.28e-154
Score: 1222.00
Percent Similarity: 100.00%
Best Local Similarity: 99.58%
Query Match: 99.76%
DB: 4
Length: 5013
Matches: 238
Conservative: 1
Mismatches: 0
Indels: 0
Gaps: 0
US-10-035-300A-2 (1-239) x US-09-183-188B-5 (1-5013)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 413 ATGGCTACCCACACATTAATGCAAGTAATGGCGATTTTCGCTGACGTAGTTTGTATGCCA 472

Qy 21 GlyAspProLeuArgAlaIleAsnAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 473 GCGACCGCGTGGTGGAGTATATTGCTGAACTTTCTTGAAGATGCCGTGAAGTG 532

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerVal 60
Db 533 AACAACTGTCGCGTATGCTGGGCTTCCACGGTACTTACAAAGCCGCCAAAATTTCCGTA 592

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAsp 80
Db 593 ATGGGTCAACGGTATGGGTATCCCGTCTGCTCATCTACACCAAGAACTGATCACCGAT 652

Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 653 TTCGGCGTGAAGAAATATATCCGGTGGTTCCTGTGGCGCAGTTCTGCGCACGTAAAA 712

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 713 CTGCGCACGTCGTATCGGTATGGGTGCGCTGCACCGATTCCAAAGTTAAACCGCATCCGT 772

Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 773 TTTAAAGACCATGACTTTGGCGTATCGCTGATCGCTGCGTAACTGTTCTCCGCTGACCTGTT 832

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 833 GCAGCTAAAGCAGCTGGGTATTGATGCTCGCGTGGGTAACTGTTCTCCGCTGACCTGTT 892

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyIleLeuGlyValGlu 180
Db 893 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAAAAATACGCGATTCCTCGCGGTGAA 952

Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 953 ATGGAAACGGCTGTTATCTACGGGCTCGCTGCAGAAATTTGGCGGAAGCCCTGCACATC 1012

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1013 TGCACCGTATCTGACACATCCGCACTACGAGCAGACCACTGCGCTGAGCGTCAGACT 1072

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 1129
```

RESULT 2

US-09-489-039A-1631
; Sequence 1631, Application US/09489039A
; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; PRIOR FILING DATE: 2000-01-27

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 1631

; LENGTH: 732

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-1631

Alignment Scores:

Pred. No.: 3.48e-147 Length: 732

Score: 1162.00 Matches: 226

Percent Similarity: 97.07% Conservative: 6

Best Local Similarity: 94.56% Mismatches: 7

Query Match: 94.86% Indels: 0

DB: 4 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-489-039A-1631 (1-732)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20

Db 13 ATGGCAACTCTCACATTAACGCAAGAAATGGGCGATTTTCGCTGACGTGCTTTGTATGCCG 72

Qy 21 GlyAspProLeuArgAlaLysTyrlsIleAlaGluThrPheLeuGluAspAlaArgGluVal 40

Db 73 GCGACCGCGTGGCGCGAAGCAATTGCAGAAACCTTCTCGAAGACGTCGCGTGAAGTG 132

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerVal 60

Db 133 AACAACTGTCGCGCGCATGCTGGGCTTTCACGGTACCTATAAAAGCCGCCAAAATCTCCGTG 192

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAsp 80

Db 193 ATGGGCGACCGGATGGGATCCGCTCTGCTCTATCTACACCAAGAGCTGATCCCGAT 252

Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100

Db 253 TTCGGCGTGAAGAAATCATCCGCTGGGCTCTCGCGCGCGGTGGGTGAAGATGTCAA 312

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120

Db 313 CTGCGTGACGTGATGCGCATGGGCGGTGCGCGCTCTAAAGTGAACCGCTGTCGT 372

Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140

Db 373 TTCAAAGACCATGATTTTGGCGGATTCGCGATTCGCGTATGCGGTAGAC 432

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160

Db 433 GCGCGAAAACGCTGGCGGTGACGCGCGCTCGGCAACATCTTCTCCGACGACCTGTT 492

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrlsGlyIleLeuGlyValGlu 180

Db 493 TATACCGCGGACCCGCTCCATGTTTCGACGTGATGGAAAAATACCGCATTCGCGCGTGGAA 552

Qy 181 MetGluAlaAlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200

Db 553 ATGGAAAGCGCGGTATCTACGCGGTGGCGCGAGTTCGGCGCGAAAGCGCTGACCATC 612

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220

Db 613 TGCACCGTCTCCGACCACTCCGTACCCACGACGACCACTGCGCGTGAACGTCAGACG 672

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239

Db 673 ACCTTCAACGACATGATCAAAATCGCGTGGAAATCCGCTGCTGCGCGGATAAAGAG 729

RESULT 3

US-09-543-681A-2371

; Sequence 2371, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 2371

; LENGTH: 795

; TYPE: DNA

; ORGANISM: Proteus mirabilis

US-09-543-681A-2371

QY 163 ProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGluMetGlu 182
DB 540396 CCCGATGTGGAAATGTCGATGTAATGGAATAATACGGCAATTTAGGTGGAAATGGAA 540337
QY 183 AlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThr 202
DB 540336 GCTGCTGGAATTTATGTTGTCGTCAGAAATATGTTGCAAAAGCATTTAAGCAATTTGTACC 540277
QY 203 ValSerAspHisIleArgThrHisGluGlnThrThraAlaGluArgGlnThrPhe 222
DB 540276 GTGTCTGACCATATTCGTACTCAGAACAAACACCGCAGAAAGCATTAATTAACATTC 540217
QY 223 AsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAsp 237
DB 540216 AATGATATGATTGGAATTTGGTTAGATTCAGTATTAAATTTGGTGTAT 540172

RESULT 5

US-09-643-990A-1/c
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Alignment Scores:

Pred. No.:	8,14e-122	Length:	1830121
Score:	1013.00	Matches:	197
Percent Similarity:	90.21%	Conservative:	15
Best Local Similarity:	83.83%	Mismatches:	23
Query Match:	82.69%	Indels:	0

DB: 4 Gaps: 0
US-10-035-300A-2 (1-239) x US-09-643-990A-1 (1-1830121)
QY 3 ThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAsp 22
DB 540876 ACTCCACATATTAAACGCGCTGAAGGTGCATTTGCTGATGTTGTTTAAATGCTCGCAT 540817
QY 23 ProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsnAsn 42
DB 540816 CCATCTTCGTCGCAAAATATATTGCAAAACCTTTTTTACAAGATGTTGTTGAAGTAACGAAT 540757
QY 43 ValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMetGly 62
DB 540756 GTTCGTAATATGCTTGGTTTACTGGAACCTTATAAGGTCTGTAATAATTTCTATTATGGGG 540697
QY 63 HisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPheGly 82
DB 540696 CACGGTATGGGATTCATCTGCTCTATTATTACCGCAAGAAATTAATCACTGAATATATGC 540637
QY 83 ValLysIleValIleAlaArgValGlySerCysGlyAlaValLeuProHisValLysLeuArg 102
DB 540636 GTGAAAAAAATTAATCCGTGTAGGTTCTTGTGGCACCTGTTGCTATGGACGTGAAGTCCGC 540577
QY 103 AspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLys 122
DB 540576 GATGTGATTATCGTTTGGTGTGCTACTGATTTCAAAGTAACCGTATTCGTTTCAAA 540517
QY 123 AspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAla 142
DB 540516 GATAACGACTTTNCAGCTATTGTCGCTTGTATATGTCACACGCGCTGTTCAAGCTGCA 540457
QY 143 LysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPheTyrSer 162
DB 540456 AAAGCAAAAGGTAAAGTCGTGCTGTAGTAATTTATTTCTCGCGGATTTATTCTATACA 540397
QY 163 ProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGluMetGlu 182
DB 540396 CCCGATGTGGAATGTTTCGATGTAATGGAAATAACGGCATTTTAGGTGTGGAATGGAA 540337
QY 183 AlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThr 202
DB 540336 GCTGCTGGAATTTATGTTGCTGTCGAGAATATGTTGCAAAAGCATTTAAGCATTTGTACC 540277
QY 203 ValSerAspHisIleArgThrHisGluGlnThrThraAlaGluArgGlnThrThrPhe 222
DB 540276 GTGTCTGACCATATTCGTACTCAGAACAAACACCGCAGAAAGCATTAATTAACATTC 540217
QY 223 AsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAsp 237
DB 540216 AATGATATGATTGGAATTTGGTTAGATTCAGTATTAAATTTGGTGTAT 540172
RESULT 6
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRO
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.

US-09-790-988-1

Alignment Scores:

Pred. No.: 9,1e-91 Length: 640681
Score: 778.00 Matches: 144
Percent Similarity: 79.91% Conservative: 43
Best Local Similarity: 61.54% Mismatches: 47
Query Match: 63.51% Indels: 0
DB: 4 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-790-988-1 (1-640681)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 575671 ATGCTACTCCACATATTAATAGTAAAGATGATGATTTTTCAGATATATGTTTAAATGCGG 575612

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 575611 GGAGACCAGTTCGTGCAAAATATATGCTGAAATATTTTAAGTAAATTTTGTTCNAGTA 575552

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyTyrThrLysGlyArgLysIleSerVal 60
Db 575551 AATGATACATCGTTTAAATGTTAGCTTATACCGGATTTTATATAAAATAGAAAGATTTCATC 575492

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 575491 ATGAGTCATGTTATAGGAATACCATCAGCTTCTCTATATACGAGAGATTAAATTATTGAA 575432

Qy 81 PheGlyValLysLysIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 575431 TTTAATGTAATAAATTTATGATAGAACTTGTGTGCTGTACGAGATGATATAAG 575372

Qy 101 LeuArgAspValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 575371 TTACGTGATATAGTAATCAGTATGGAGCTTCTACTGATTCTAAGTAATAAGATAAGA 575312

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 575311 TTTAATGATCATGACTTTGCTGCCATTGCAGATTTTGATATGATGATTAATATATATCTCA 575252

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 575251 ATTTCAAAAAAATGAAATAAAGTCTCCATGTTGTAATTTTTTACAAACAGATCTTTT 575192

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 575191 TATAATGATGATAAAAAATGCTAAATATTTAAAAAATATATATATTATTGGAGTTGAC 575132

Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 575131 ATGGAGACTGGCGGAATATATGGAGTGGCTTCTGAATTTAAAGTTTCAAGCATTTATCGATA 575072

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 575071 TGTACAGTATCTGATCATATACAAATAAAGATTCTTTTCATCAAGAAAGAGATCA 575012

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeu 234
Db 575011 AGTTTAAATGATATGATGAAGTACCTTGGAACTCTGTTTA 574970

RESULT 7

US-09-902-540-3568
; Sequence 3568, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10

; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 3568
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3568

Alignment Scores:

Pred. No.: 1,55e-88 Length: 744
Score: 728.00 Matches: 139
Percent Similarity: 77.49% Conservative: 52
Best Local Similarity: 60.17% Mismatches: 52
Query Match: 59.43% Indels: 0
DB: 4 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-902-540-3568 (1-744)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCGACTCCTCATATCTCCGCTGCCCTCGCTGACTTCGCTGACGTGGTCTCATGCT 60

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GCGACCCGCTCCGCGCTCGTTATCTCCAGCCGCTTCTGGAGGTGCCCGCGAGGTC 120

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyTyrLysGlyArgLysIleSerVal 60
Db 121 ACTTCGTCGCAACATGCTCGGCTTCACCGGAGCTTCCGGGCGCGGCGGTGCGGTG 180

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 181 ATGGGCGACGCGATGGCGTTCCTCCATCTCCATCTACGCCACCGAGCTCATCAAGACG 240

Qy 81 PheGlyValLysLysIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TATGGGTCGCGTGATCATCCGCTGGCGAGCTCGCGCGCTGAGCACCGACGTGAAG 300

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 GTCGCGGAGTTCATGTCGCGACCGGGCGCGCGACGAGCTCAACGTGAATCGGATCGG 360

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 CTGATGGCGATGACTTCGCGCGGTGGCGACTTCACGCTCGCGCGCGCGGCGCATGGAG 420

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCGCGGAGCGCGCAACAGCTGTGCGCGCGCGCGCTTCTTACCTCCGACCTCTTC 480

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACCACCGCAGGAGCAGCTCAACGCCACCTTCGCGCGGATGGGTGTCTGCGCGCTCGAG 540

Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAGTGGCGCGCTCTACGGGTGGCGCGGAGTGGCGCGCGCGCGCGCTGGGGCTG 600

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 CTCACGGTGTGCGACCATCATCACCGGGGAGAGTCTTCACGCGCGAGGAGCGGAGACG 660

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGlu 231
Db 661 ACGTTCGACGAGATGATGAATCTCGCCTCGAC 693

RESULT 8

US-09-902-540-1171/c
; Sequence 1171, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.

```
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1171
; LENGTH: 19019
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(19019)
; OTHER INFORMATION: unsure at all n locations
US-09-902-540-1171

Alignment Scores:
Pred. No.: 2,29e-86 Length: 19019
Score: 728.00 Matches: 139
Percent Similarity: 77.49% Conservative: 40
Best Local Similarity: 60.17% Mismatches: 52
Query Match: 59.43% Indels: 0
DB: 4 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-902-540-1171 (1-19019)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 16556 ATGGCGACTCTCATATCTCGCTGCCCTGGTGACTTCGCTGACGTGGTCTCATCGCT 16497

Qy 21 GlyAspProLeuArgAlaLysTyriLeAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 16496 GGCACCGCTCCGCGTCTGTTACATCTCCGACCGCTTCTTGGAAAGTGGCCGCGAGGTC 16437

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyriLysGlyArgLysIleSerVal 60
Db 16436 ACCCTCCGTGGCAACATGCTCGGCTTCCCGGGACCTTCCGGGGCCGCGGTGTCGGTG 16377

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyriThrLysGluLeuIleThrAsp 80
Db 16376 ATGGGCGACGCGATGGGGTTCCTTCATCTACGCCCGAGCTCATCAAGC 16317

Qy 81 PheGlyValLysLysIlelleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 16316 TATGGGTGGCGTGATCATCCGCTGGCGAGCTCGCGCGCTGACGACCGACGTGAAG 16257

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 16256 GTCCGGAGGTGCATCTGTCGGACCGGGGCGCGGACGACTCCACGCTGAATCGGATCGG 16197

Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 16196 CTGATGGGCATGACTTCGCGCGGTGGCGGACTTCACGCTGCGCGCGCGGCGCATGGAG 16137

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 16136 GCGCGGAGCGGCGCAACAGCCTGTGCGCGCGCGCGCGCTTTCACCTCCGACCTCTTC 16077

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyriGlyIleLeuGlyValGlu 180
Db 16076 TACCACCGCAGGAGCAGCTCAACGCCACCTTGGCGGGATGGGTGTCTTGGCGCTCGAG 16017

Qy 181 MetGluAlaAlaGlyIleTyriGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 16016 ATGAGGTGCGCGGCTCTACGGCGTGGCGCGAGTCGGCGCGCGCGCTGGGCGTG 15957

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 15956 CTCACGGTGTGGACCATCATCACCAGGAGAGTCTCACCGCGGAGAGTCTCACCGCGGAGGCGGACG 15897
```

```
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGlu 231
Db 15896 ACGTTCAGCAGATGATTGAATCGCCCTGGAC 15864
```

```
RESULT 9
US-09-107-532A-1185
; Sequence 1185, Application US/09107532A
; Patent No. 658275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1185:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...735
; SEQUENCE DESCRIPTION: SEQ ID NO: 1185:
US-09-107-532A-1185
```

```
Alignment Scores: 3,39e-72 Length: 735
Pred. No.: 607.00 Matches: 122
Score: 70.13% Conservative: 40
Percent Similarity: 52.81% Mismatches: 67
Best Local Similarity: 49.55% Indels: 2
Query Match: 4 Gaps: 2
DB:
```

```
US-10-035-300A-2 (1-239) x US-09-107-532A-1185 (1-735)

Qy 5 HistLeaAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAspProLeu 24
Db 37 CATATTGACGGAAGAGAGCGGATCGCCGATAGATTTTACTTCGGGAGATCCTTTG 96
Qy 25 ArgAlaLysTyriLeAlaGluThrPheLeuGluAspAlaArgGluValAsnValArg 44
```

Db 97 CGAGCTAATATATGCGGAGACTTTTGAAGATCCAGTTTGTACATCAGGTACGC 156
Qy 45 GlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysGlyIleSerValMetGlyHisGly 64
Db 157 GGTATGTTAGGCTTTACCGCAATATAAGGAGAACGTATTTCTGTTCAAGGAACAGGA 216
Qy 65 ValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAspPheGlyVallys 84
Db 217 ATGGGAATGCCATCAGCCACCATCTATCCCATGAATTTGATCCAGTCGTATGGTGAAG 276
Qy 85 LysIleIleArgValGlySerCysGlyValaValLeuProHisValLysLeuArgAspVal 104
Db 277 AAATGATTCAGTAGGTACATCGCGTCATGTCMAAAGACGTCATGTCGTGATTTA 336
Qy 105 VallIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLysAspHis 124
Db 337 GTACTAGCACAGGTGCGAGCCAGCTTCTTCGATGATCGAAGAAATTTCCAGCCCTT 396
Qy 125 AspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAlaLysAla 144
Db 397 CATTTTCTCCATCAGTGATTTCAATCTATTACTGAAAGCATATGATAGCTAAAGAA 456
Qy 145 LeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPheTyrlsSerProAsp 164
Db 457 AAGGCTATACGGTTCATGTAGGAATGTTCTTCGGAAGATTTCTTCTCAAAAGATGAT 516
Qy 165 ---GlyGluMetPheAspValMetGlyLysTyrlsGlyIleLeuGlyValGluMetGluAla 183
Db 517 TTGACAGAGAGCTTCCAATTAGCTGAA---CTTGGCGTTTATGAGTGGAAATGGAACGA 573
Qy 184 AlaGlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrVal 203
Db 574 GCGGCTCTTTATTAATCTAGGCGCGAAATACCATGTGCAGACATTAAGTTTATGACCGTT 633
Qy 204 SerAspHisIleArgThrHisGluIleThrThrAlaAlaGluArgGlnThrThrPheAsn 223
Db 634 AGCACCATTGATACCGGGAAGAAACACAGCTGCAGAACGTCAATCTACGTTCAAT 693
Qy 224 AspMetIleLysIleAlaLeuGluSerValLeu 234
Db 694 GAATGATCGAAGTAGGATTGGAAACACGCGATA 726

RESULT 10

US-09-134-000C-2608
; Sequence 2608, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055, 778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2608
; LENGTH: 714
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2608

Alignment Scores:
Pred. No.: 4,71e-70 Length: 714
Score: 591.00 Matches: 117
Percent Similarity: 69.13% Conservative: 42
Best Local Similarity: 50.87% Mismatches: 71
Query Match: 48.24% Indels: 0
DB: 4 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-134-000C-2608 (1-714)

Qy 5 HisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAspProLeu 24
Db 16 CATATCGAAGCAAAACCCAGGTGAATTCGCCGATAAAATTTTACTACAGGGGATCCCTC 75
Qy 25 ArgAlaLysTyrlsIleAlaGluThrPheLeuGluAspAlaArgGluValAsnValArg 44
Db 76 CGACAAAATATATATGCGAAACATTTTATAGAAATCTCTGCTCTATAATCAAGTTTCT 135
Qy 45 GlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerValMetGlyHisGly 64
Db 136 GGGATGTTAGTTACACAGGACGCTATAAGGACAGCGGTGTTTCAGTCCCAAGGAACAGG 195
Qy 65 ValGlyIleProSerCysSerIleTyrlsGlyLeuLeuIleThrAspPheGlyVallys 84
Db 196 ATGGGATGCGCTTCAGCAGGATTTATATGCCCATGAGTTAATACTCTTACGATGTGAAG 255
Qy 85 LysIleIleArgValGlySerCysGlyValaValLeuProHisValLysLeuArgAspVal 104
Db 256 AAGTTAATTCGTGTGAACGTTGTTCTATTATCAGAAAAGTCAACGTTTCGTGAATTA 315
Qy 105 VallIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLysAspHis 124
Db 316 GTTATCGCACAGCAGCAGCGACGCTTTTCCCGGATTCGCAATGATTTTCTTAATAT 375
Qy 125 AspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAlaLysAla 144
Db 376 GATTTTCCACAAATTCGAGCTTTGATCTTTTACTCAAACTTACGAAATTTGCTAAAGCA 435
Qy 145 LeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPheTyrlsSerProAsp 164
Db 436 AAAGGCTTTTACTACCCATGTTGTAACGTTATGTCAGATGATGTTTTTACAAAGATAGT 495
Qy 165 GlyGluMetPheAspValMetGlyLysTyrlsGlyIleLeuGlyValGluMetGluAla 184
Db 496 TTAGTATGTCCTTTCGTTTAGTTCGATACGGGTATTTAGCGATTGAATGGAGCGCA 555
Qy 185 GlyIleTyrlsGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrValser 204
Db 556 ATGTTATATTATTATGACGCAAAATTTGATGTAGAACATTTGGCGATTATGCGGTGAGT 615
Qy 205 AspHisIleArgThrHisGluIleThrThrAlaAlaGluArgGlnThrThrPheAsnAsp 224
Db 616 GATAGCTTGTGACAGCGCAAGAAACACGCGCAGAAACGCAACGCGATTTAATGAA 675
Qy 225 MetIleLysIleAlaLeuGluSerValLeu 234
Db 676 ATGATTCGAAGTTGTTAGAAACCGGCAATT 705

RESULT 11

US-09-583-110-37
; Sequence 37, Application US/09583110
; Patent No. 6698703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 37
; LENGTH: 711
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-37

```
Alignment Scores:
Pred. No.: 4,26e-68 Length: 711
Score: 576.50 Matches: 115
Percent Similarity: 68.83% Conservative: 44
Best Local Similarity: 49.78% Mismatches: 71
Query Match: 47.06% Indels: 1
DB: 4 Gaps: 1

US-10-035-300A-2 (1-239) x US-09-583-110-37 (1-711)

Qy 5 HisileAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAspProLeu 24
Db 10 CATATTGCTGCTCAGCAGGCGTAAATGCTGATAAAATCTTCTCTCTGGGATCCTCTT 69
Qy 25 ArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsnAsnValArg 44
Db 70 CGTGCTAAGTTTATTGCGGAGAAATTTCTTGGTGATGCTGTTGTTTAAACGAAGTGGT 129
Qy 45 GlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMetGlyHisGly 64
Db 130 AACATGTTTGGTTACACTGCTGACTTACAGGGTCACCGTGATCTGTCTGCGAAGTGGG 189
Qy 65 ValGlyIleProSerCysSerIleTyrThrLysGlyLeuLeuThrAspPheGlyValLys 84
Db 190 ATGGGAATGCCATCTATTTCGATTATATCGCGTGAGTTAATCGTAGACTACGGTGAAG 249
Qy 85 LysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLeuArgAspVal 104
Db 250 AAATTGATTGCTGGGAACATCGCAGGTTCTTTGAATGAAGAGGTTTCATGTCGTGAATTA 309
Qy 105 ValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLysAspHis 124
Db 310 GTTTGGCGCAGCGGCTGCAACCACTCAACATCGTTGTAATGCTGGCCACAGTAC 369
Qy 125 AspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAlaLysAla 144
Db 370 GATTTTCCAAATTTGCTAGCTTTGATTGTTGATAAAGCCTACCATATCGCCMAAGAA 429
Qy 145 LeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPheTyrSerProAsp 164
Db 430 CTTGGTATGACTACTCAGCTTGGGAAGCTTTTGTGTCATCTGCTTTTACTCAAAATAC 489
Qy 165 GlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrValSer 204
Db 490 TTTGAAAGAATATCGAGCTTGGTAATGGGAGTCAAGGCTGTGGAATGGAGAGCA 549
Qy 185 GlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrValSer 204
Db 550 GCTCTTTACTATCTTGTGCTGCAATACCATGTTGATGCGCTAGCTATCATGACCATCT 609
Qy 205 AspHis---IleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThrThrPheAsn 223
Db 610 GATAGCTTGGTCAATCCAGACGAGACACACATGCGAAGAAGCTCAAAATACCTTCACT 669
Qy 224 AspMetIleLysIleAlaLeuGluSerValLeu 234
Db 670 GATATGATGAAGGTTGGTTTGGAAACCTTGATT 702

RESULT 12
US-09-107-433-1967
Sequence 1967, Application US/09107433
Patent No. 6800744
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNO
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
```

Qy 125 AspPheAlaAlaIleAlaAaspPheAaspMetValArgAenAlaValAaspAlaAlaLysAla 144
Db 373 GATTTTCCACAAATGCTAGCTTTGATTTGCTGATTAAGCTACCATATGCGCAAGAA 432
Qy 145 LeuGlyIleAaspAlaArgValGlyAenLeuPheSerAlaAaspLeuPheTyrSerProAasp 164
Db 433 CTTGGTATGACTACTCAGCTTGGGAACGTTTGTCTGATCTGATGATCTTTTACTCAAAATAC 492
Qy 165 GlyGluMetPheAaspValMetGluLysTyrGlyIleLeuGlyValGluMetGluAlaAla 184
Db 493 TTTGAAAGAAATATCGAGCTTGGTAAATGGGAGTCAAGGCTGTGGAAATGGAAGCAGCA 552
Qy 185 GlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrValSer 204
Db 553 GCTCTTTACTACTTGTCTGCGCAATACCATGTTGATGCGCTAGCTATCATGACCATCTCT 612
Qy 205 AspHis---IleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThrThrPheAen 223
Db 613 GATAGCTTGGTCAATCCAGACGAACACAACTGCAGAAAGAACTGCAAAATATACCTTCTACT 672
Qy 224 AspMetIleLysIleAlaLeuGluSerValLeu 234
Db 673 GATATGATGAAGGTGGTGGTTCGAAACCTTGATT 705

RESULT 13

US-08-961-527-128
; Sequence 128, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 128:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

US-08-961-527-128
Alignment Scores:
Pred. No.: 5,39e-66 Length: 13440
Score: 575.50 Matches: 115
Percent Similarity: 68.83% Conservative: 44
Best Local Similarity: 49.78% Mismatches: 71
Query Match: 46.98% Indels: 1
DB: 3 Gaps: 1

US-10-035-300A-2 (1-239) x US-08-961-527-128 (1-13440)
Qy 5 HisIleAenAlaGluMetGlyAaspPheAlaAaspValValLeuMetProGlyAaspProLeu 24
Db 10453 CATATTCTCTCAGCAGGAGTGAATTCGTGATAAAATTCCTCTCTCTGGGATCCCTCT 10512
Qy 25 ArgAlaLysTyrIleAlaGluThrPheLeuGluAaspAlaArgGluValAenAaspValArg 44
Db 10513 CGTGCTAAGTTTATTGCGGAGAAATTCCTTGATGATGCTGTTGTTTAAACGAAGTGGT 10572
Qy 45 GlyMetLeuGlyPheThrGlyThrTyrGlyArgLysIleSerValMetGlyHisGly 64
Db 10573 AACATGTTGGTATACACTGCTACTTCAAGGGTCACTGTGTATCTGATCTGATCGGAACCTGGG 10632
Qy 65 ValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAaspGlyValLys 84
Db 10633 ATGGGAATGCACUATATTCATTTATGCGGTGAGTTAATCGTAGACTAGCGGTGTGAG 10692
Qy 85 LysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLeuArgAaspVal 104
Db 10693 AAATTGATTCGTGTGGAACTGCAGGTTCTTTGAATGAAGAGGTTTCATGTTGCGTAATTA 10752
Qy 105 ValIleGlyMetGlyAlaCysThrAaspSerLysValAenArgIleArgPheLysAaspHis 124
Db 10753 GTTTTGGCGAGCGGCTGCAACCACTCAACATCGTTCGTAATGACTGGCCACAGTAC 10812
Qy 125 AspPheAlaAlaIleAlaAaspPheAaspMetValArgAenAlaValAaspAlaAlaLysAla 144
Db 10813 GATTTTCCAAATTTGCTAGCTTTGATTTGCTGATAAAGCTACCATATGCGCAAAATA 10872
Qy 145 LeuGlyIleAaspAlaArgValGlyAenLeuPheSerAlaAaspLeuPheTyrSerProAasp 164
Db 10873 CTTGGTATGACTACTCAGCTTGGGAACGTTTGTGTCATCTGATGCTTTTACTCAAAATAC 10932
Qy 165 GlyGluMetPheAaspValMetGluLysTyrGlyIleLeuGlyValGluMetGluAlaAla 184
Db 10933 TTTGAAAGAAATATCGAGCTTGGTAAATGGGAGTCAAGGCTGTGGAATGGAAGCAGCA 10992
Qy 185 GlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIleCysThrValSer 204
Db 10993 GCTCTTACTATCTTGTCTGCCAATACCATGTTGATGCGGTAGCTATCATGACCATCTCT 11052
Qy 205 AspHis---IleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThrThrPheAen 223
Db 11053 GATAGCTTGGTCAATCCAGACGAACACAACTGCAGAAAGAACTGCAAAATATACCTTCACT 11112
Qy 224 AspMetIleLysIleAlaLeuGluSerValLeu 234
Db 11113 GATATGATGAAGGTGGTGGTTCGAAACCTTGATT 11145

RESULT 14

US-09-134-001C-1402
; Sequence 1402, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1402
; LENGTH: 732
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1402

```
Alignment Scores:
Pred. No.: 1.81e-67 Length: 732
Score: 572.00 Matches: 123
Percent Similarity: 66.23% Conservative: 30
Best Local Similarity: 53.25% Mismatches: 76
Query Match: 46.69% Indels: 2
DB: 3 Gaps: 2

US-10-035-300A-2 (1-239) x US-09-134-001C-1402 (1-732)

Qy 2 AlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGly 21
Db 31 AGTACACCTCACATTAAACCCAAATGGAGTTAAATAGCTAAATACTGTTAATGCCAGGC 90
Qy 22 AspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsn 41
Db 91 GATCCATTACGTCTCAATATATTGCTGAAATTTTGTAGAAAATTTTGAACAAATGTTGAACAAATCAAT 150
Qy 42 AsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMet 61
Db 151 ACAGTAGCTAATATGTTGGTTACACAGGAACCTTATAAGGCCAAAGATTTCTGTGATG 210
Qy 62 GlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPhe 81
Db 211 GGCTCTGGGATGGGAATTCCAAGTATTGTTATTTATCTTATGAACCTTTACCATTTCTTT 270
Qy 82 GlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLeu 101
Db 271 GATGTGGATACAATCATTCGCGTAGGTTCATGTGGCGCTCTTCAAGGAAGATGTCAAATTTA 330
Qy 102 ArgAspValValIleGlyMetGlyAlaCysThrAspSerLys---ValAsnArgIleArg 120
Db 331 TATGATGTGATTATTGACACAGCGCTCCACAACTCAAAATATGTTGATCAATTTAAT 390
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 391 ATACCGGGTCAC---TTTGGCGCTATTGCTGATTTAATCTAGTAGCTAAAGCTAAGAAA 447
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 448 GCTGCTGATGAATTTGGTGCATATCATGTAGGTAAACGTTTGTTCATCTGATACATTT 507
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 508 TATAATCTGNTCAACTTTTAATGATTTCATGGAAGAAAGATGGTATCTTAGGTATCGAA 567
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 568 ATGGAATCTCGAGTCTATATTTAAATGCCATCCATGCCAATAAAAGGCTTTTAGGTATT 627
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrAlaAlaGluArgGlnThr 220
Db 628 TTTACAGTAGTGATCATATTTCTAAGAGATGAAGCCACTAGTCAGAAGAGAGACAAACA 687
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGlu 231
Db 688 TCATTTACTCAATGATGGAATAGCGCTCGAA 720

RESULT 15
US-09-710-279-3511
; Sequence 3511, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3511
```

```
; LENGTH: 2962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3511

Alignment Scores:
Pred. No.: 3.96e-66 Length: 2962
Score: 569.00 Matches: 122
Percent Similarity: 66.23% Conservative: 31
Best Local Similarity: 52.81% Mismatches: 76
Query Match: 46.45% Indels: 2
DB: 4 Gaps: 2

US-10-035-300A-2 (1-239) x US-09-710-279-3511 (1-2962)

Qy 2 AlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGly 21
Db 1638 AGTACACCTCACATTAAACCCAAATGGAGTTAAATAGCTAAATACTGTTAATGCCAGGC 1697
Qy 22 AspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsn 41
Db 1698 GACCCATTACGTCTCAATATATTGCTGAAATTTTGTAGAAAATTTTGAACAAATGTTCAAT 1757
Qy 42 AsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMet 61
Db 1758 ACAGTAGCTAATATGTTGGTTACACAGGAACCTTATAAGGCCAAAGATTTCTGTGATG 1817
Qy 62 GlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPhe 81
Db 1818 GGCTCTGGGATGGGAATTCCAAGTATTGTTATTTATCTTATGAACCTTTACCATTTCTTT 1877
Qy 82 GlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLysLeu 101
Db 1878 GATGTGGATACAATCATTCGCGTAGGTTCATGTGGCGCTCTTCAAGGAAGATGTCAAATTTA 1937
Qy 102 ArgAspValValIleGlyMetGlyAlaCysThrAspSerLys---ValAsnArgIleArg 120
Db 1938 TATGATGTGATTATTGACACAGCGCTTCACAACTCCAAATATGTTGATCAATTTAAT 1997
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 1998 ATACCGGGTCAC---TTTGGCGCTATTGCTGATTTAATCTAGTAGCTAAAGCTAAGAAA 2054
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 2055 GCTGCTGATGAATTTGGTGCATATCATGTAGGTAAACGTTTGTGTCATCTGATACATTT 2114
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 2115 TATAATGCTGATTCAACTTTTAAATGATTTCATGGAAGAAAGATGGGTATCTTAGGTATCGAA 2174
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 2175 ATGGAATCTCGAGTCTATATTTAAATGCCATCCATGCCAATAAAAGGCTTTTAGGTATT 2234
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrAlaAlaGluArgGlnThr 220
Db 2235 TTTACAGTAGTGATCATATTTCTAAGAGATGAAGCCACTAGTCAGAAGAGAGACAAACA 2294
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGlu 231
Db 2295 TCATTTACTCAATGATGGAATAGCGCTCGAA 2327

Search completed: July 23, 2005, 04:45:33
Job time : 1196 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 23, 2005, 03:20:48 ; Search time 742 Seconds
(without alignments)
2043.501 Million cell updates/sec
Title: US-10-035-300A-2
Perfect score: 1225
Sequence: 1 MATPHINAMGDFADVVLMP.....TTFNDMIKIALESVLLGXDE 239

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 7173243 seqs, 3172129809 residues

Total number of hits satisfying chosen parameters: 14346486

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10035300/runat_18072005_165352_28093/app_query.fasta_1.391
-DB=Published Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=Blosum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10035300@cgn1.1.723 @runat_18072005_165352_28093
-NCPU=6 -ICPU=3 -NO WMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:**

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq.*
- 24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1225	100.0	720	17	US-10-035-300A-1
2	1222	99.8	720	21	US-10-035-300A-3
3	1222	99.8	3031	10	US-09-891-865A-12
4	1222	99.8	3031	24	US-11-036-497-12
5	1222	99.8	3128	10	US-09-891-865A-13
6	1222	99.8	3128	24	US-11-036-497-13
7	1222	99.8	3383	10	US-09-891-865A-3
8	1222	99.8	3383	24	US-11-036-497-3
9	1222	99.8	3934	10	US-09-891-865A-14
10	1222	99.8	3934	24	US-11-036-497-14
11	1222	99.8	4189	10	US-09-891-865A-5
12	1222	99.8	4189	24	US-11-036-497-5
13	1222	99.8	5013	14	US-10-198-034-5
14	1222	99.8	5013	15	US-10-198-228-5
15	1222	99.8	5241	10	US-09-891-865A-7
16	1222	99.8	5241	24	US-11-036-497-7
17	1222	99.8	5495	10	US-09-891-865A-4
18	1222	99.8	5495	24	US-11-036-497-4
19	1222	99.8	6046	10	US-09-891-865A-15
20	1222	99.8	6046	24	US-11-036-497-15
21	1222	99.8	6269	10	US-09-891-865A-9
22	1222	99.8	6269	24	US-11-036-497-9
23	1222	99.8	6299	10	US-09-891-865A-10
24	1222	99.8	6299	24	US-11-036-497-10
25	1222	99.8	6301	10	US-09-891-865A-6
26	1222	99.8	6301	24	US-11-036-497-6
27	1218	99.4	720	17	US-10-035-300A-3
28	1013	82.7	1830121	17	US-10-329-670-1
29	1013	82.7	1830121	20	US-10-158-865-1
30	1013	82.7	1830121	22	US-10-981-988-1
31	778	63.5	640681	9	US-09-730-988-1
32	700	57.1	5822	10	US-09-891-865A-8
33	700	57.1	5822	24	US-11-036-497-8
34	693.5	56.6	899	19	US-10-772-272A-15
35	675	55.1	843	9	US-09-895-913A-147
36	671.5	54.8	319630	17	US-10-398-221-7
37	671.5	54.8	3011208	17	US-10-398-221-2058
38	670.5	54.7	1171	17	US-10-398-221-3116
39	663	54.1	693	18	US-10-335-977-3426
40	601	49.1	16484	9	US-09-070-927A-158
41	593	48.4	687	18	US-10-335-977-3425
42	576.5	47.1	714	22	US-10-617-320-1967
43	575.5	47.0	708	21	US-10-472-928-1597
44	575.5	47.0	13440	8	US-08-961-527-128
45	575.5	47.0	13440	17	US-10-158-844-128

ALIGNMENTS

RESULT 1
US-10-035-300A-1
; Sequence 1, Application US/10035300A
; Publication No. US20030228576A1
; GENERAL INFORMATION:
; APPLICANT: Ealick, Steven E.
; APPLICANT: Parker, William B.
; APPLICANT: Secret, John A.
; APPLICANT: Sorscher, Eric J.
; TITLE OF INVENTION: MUTANT PURINE NUCLEOSIDE PHOSPHORYLASE PROTEINS AND CELLULAR DELIV
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: UAB-20702/22
; CURRENT APPLICATION NUMBER: US/10/035.300A
; CURRENT FILING DATE: 2001-10-26
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-035-300A-1

Alignment Scores:
Pred. No.: 1,96e-143 Length: 720
Score: 1225.00 Matches: 239
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-035-300A-2 (1-239) x US-10-035-300A-1 (1-720)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGTCGACGTAGTTTTGATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GGCACCCGCTGGTGGCGAAGTATATTCCTGAACATTTTCCTGAAGATGCCCGTGAAGT 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 121 AACAACTGTCGGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 181 ATGGCTCACGGTGTGGTATCCGCTCCGCTCCATCTACACCAAGAACTGATCACCGAT 240
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCCTGTGGCGAGTTCTGCCGACGTAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGGCGACGCTGTTATCGGTATGGGTGCTCGACCCGATTCCTCAAAAGTTAACCGCATCCGT 360
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGGCGTATCGCTGACTTCGACATGGTGGCTAAACGCGAGTAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGCAGCTGGGTATGTATCGGTGGGTAACTTCTCCGCTGACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCGGACGGCGAAATGTTTCGACGTGATGGGAAATACGGCATTTCTCGGCTGGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAAGCGGTGGTATCTACCGGCTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCAATCCGACATCAGAGCAGACCATCTGCGCTGAGCGTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 717
```

RESULT 2

US-10-891-096-3
; Sequence 3, Application US/10891096
; Publication No. US2005074857A1
; GENERAL INFORMATION:
; APPLICANT: Araki, Tadashi
; APPLICANT: Miyake, Hitoki
; APPLICANT: Okawa, Toshihiro
; TITLE OF INVENTION: Method for Producing a Pyrimidine Nucleoside Compound

; TITLE OF INVENTION: and a New Pyrimidine Nucleoside Compound
; FILE REFERENCE: 018765-161
; CURRENT APPLICATION NUMBER: US/10/891,096
; CURRENT FILING DATE: 2004-07-15
; PRIOR APPLICATION NUMBER: JP 2003-199175
; PRIOR FILING DATE: 2003-07-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-891-096-3

Alignment Scores:
Pred. No.: 4,66e-143 Length: 720
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 21 Gaps: 0

US-10-035-300A-2 (1-239) x US-10-891-096-3 (1-720)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGTCGACGTAGTTTTGATGCCA 60
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 61 GGCACCCGCTGGTGGCGAAGTATATTCCTGAACATTTTCCTGAAGATGCCCGTGAAGT 120
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 121 AACAACTGTCGGGTATGCTGGGCTTCACCGGTACTTACAAAGGCCGCAAAATTTCCGTA 180
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 181 ATGGCTCACGGTATGGGTATCCGCTCCGCTCCATCTACACCAAGAACTGATCACCGAT 240
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 241 TTCGGCGTGAAGAAATTAATCCGCGTGGGTTCCTGTGGCGAGTTCTGCCGACGTAA 300
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 301 CTGGCGACGCTGTTATCGGTATGGGTGCTCGACCCGATTCCTCAAAAGTTAACCGCATCCGT 360
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 361 TTTAAAGACCATGACTTTGGCGTATCGCTGACTTCGACATGGTGGCTAAACGCGAGTAGAT 420
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 421 GCAGCTAAAGCAGCTGGGTATGTATCGGTGGGTAACTTCTCCGCTGACCTGTTTC 480
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 481 TACTCTCGGACGGCGAAATGTTTCGACGTGATGGGAAATACGGCATTTCTCGGCTGGAA 540
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 541 ATGGAAGCGGTGGTATCTACCGGCTCGCTGCAGAAATTTGGCGGAAAGCCCTGACCATC 600
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 601 TGCACCGTATCTGACCAATCCGACATCAGAGCAGACCATCTGCGCTGAGCGTCAGACT 660
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 661 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 717
```

RESULT 3

US-09-891-865A-12
; Sequence 12, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891.865A
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 12
; LENGTH: 3031
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: udp and deoD
; OTHER INFORMATION: cloned into pGM746 without upstream ptac promoter
US-09-891-865A-12

Alignment Scores:
Pred. No.: 3.4e-142 Length: 3031
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-891-865A-12 (1-3031)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 101 ATGGCTACCCACACATTAATGCAGAAATGGCGAATTCGCTGACGTAGTTTGTATGCCA 160
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 161 GCGACCCGCTGCGTGGCGAGTATATGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 220
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 221 AACACGTTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGTA 280
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 281 ATGGCTACCGTATGGGTATCCCTGCTCCATCTACACCAAGAACTGATCACCAGAT 340
Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 341 TTCGGCGTGAAGAAATATATCCGCGTGGGTTCCTGCGCAGTTCTGCGCACGTAAAA 400
Qy 101 LeuArgAspValValIleGlyMetGlyAlaGluThrPheLeuGluAspAlaArgGluVal 120
Db 401 CTGCGCACGCTGCTATGCTGGGTATGGGTGCTTCCGCTGCTTCCAAAGTTAACCCGATCCGT 460
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheMetValArgAsnAlaValAsp 140
Db 461 TTTAAGACCATGACTTTGCGGCTATCGCTGACTTCGACATGGTTCGCTAAGCAGTAGAT 520
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 521 GCAGCTAAAGCAGTGGGTATGATGCTCGCGTGGGTAACTGTTCTCGGCTGACCTGTC 580
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 581 TACTCTCCGACGCGCAAAATGTTCCAGCTGATGAAATAATACGGCATTTCTCGGCGTGAA 640
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 641 ATGGAAAGCGGCTGTATCTACGGGCTGCTGCAGAAATTTGGCGCAAGACCTGACCATC 700
```

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 701 TGCACCGTATCTGACCACTCCGCACTCAGGAGCAGACCACTGCGCTGAGCGTCAGACT 760
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 761 ACCTTCAACGACATGATCAAAATCGCACTCGAATCCGTTCTGCTGGGGATAAAGAG 817
RESULT 4
US-11-036-497-12
; Sequence 12, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Beatezzi, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; TITLE OF INVENTION: and modified analogues thereof
; FILE REFERENCE: 02901/000J410-USO
; CURRENT APPLICATION NUMBER: US/11/036.497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 3031
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: udp and deoD cloned into pGM746 without upstream ptac promoter
US-11-036-497-12

Alignment Scores:
Pred. No.: 3.4e-142 Length: 3031
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 24 Gaps: 0

US-10-035-300A-2 (1-239) x US-11-036-497-12 (1-3031)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 101 ATGGCTACCCACACATTAATGCAGAAATGGCGAATTCGCTGACGTAGTTTGTATGCCA 160
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 161 GCGACCCGCTGCGTGGCGAGTATATGCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 220
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 221 AACACGTTTCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCCGCAAAATTTCCGTA 280
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 281 ATGGCTACCGTATGGGTATCCGCTGCTTCCATCTACACCAAGAACTGATCACCAGAT 340
Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 341 TTCGGCGTGAAGAAATATATCCGCGTGGGTTCCTGTCGCGCAGTTCTGCGCACGTAAAA 400
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 401 CTGCGCACGCTGCTATCGGTATGGGTGCTGACCGCATTTCCAAAGTTAACCCGATCCGT 460
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheMetValArgAsnAlaValAsp 140
```



```
Db 198 ATGGCTACCCACACATTAATGCGAGAAATGGCGATTTTCGCTGCGTAGTATTTCATGCCA 257
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GCGACCGCGCTGCGTGGAGTATATGCTGAAACTTTCCTTGAAGATGCCGTAAGTG 317
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 318 AACAACTGTCGCGTATGCTGGGCTTACCGGTATCTTACAAAGCGCGCAAAATTTCCGTA 377
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGlyLeuLeuThrAsp 80
Db 378 ATGGGTACCGGTATGGGTATCCCGTCCCTGCTCCATCTACCAAGAACTGATCACCGAT 437
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCGGCGTGAAGAAATATATCCGGTGGTTCCTGTGGCGAGTTCTCGCGCACGTAAGA 497
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTGGCGACGCTGCTATCGGTATGGGTGCTGCTGACCGATTCCTGCGTAAAGTTAACCGCATCCGT 557
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGATCTTTCGCGCTATCGCTGACTTCGACATGGTGGCGTAAGCGATAGAT 617
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCACCTAAAGCAGCTGGGTATGATGCTCGCTGGGTAAACCTGTTCTCGCTGACCTGTTT 677
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGlyLysTyrGlyIleLeuGlyValGlu 180
Db 678 TACTCTCGGACGCGGAAATGTTTCGACGTATGATGCTCGCTGGGTAAACCTGTTCTCGCTGACCTGTTT 737
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 738 ATGAAGCGGCTGGTATCTACGCGCTCGCTGACAGATTTGGCGGAAAGCCCTGACCATC 797
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 798 TGCACCGTATCTGACCATCTCCGACTCAGCAGCAGACCATCTGCGCTGAGCGTCAGACT 857
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 858 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAG 914
```

RESULT 7

```
US-09-891-865A-3
; Sequence 3, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: natural nucleosides and modified analogues thereof
; CURRENT APPLICATION NUMBER: US/09/891.865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: M198A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3383
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; NAME/KEY: gene
; LOCATION: (231)..(960)
; OTHER INFORMATION: deod
US-09-891-865A-3
```

Alignment Scores:

```
Pred. No.: 3,96e-142 Length: 3383
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0
```

US-10-035-300A-2 (1-239) x US-09-891-865A-3 (1-3383)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 240 ATGGCTACCCACACATTAATGCGAATATGGCGATTTTCGCTGACGTAGTATTTCATGCCA 299
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 300 GCGACCGCGCTGCGTGGAGTATATTTCTTGAAGATGCCCGTGAAGTG 359
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 360 AACAACTGTCGCGTATGCTGGGCTTACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 419
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGlyLeuLeuThrAsp 80
Db 420 ATGGGTACCGGTATGGGTATCCCGTCTCTCATCTACCAAGAACTGATCACCGAT 479
Qy 81 PheGlyValLysLysIleAlaArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 480 TTCGGCGTGAAGAAATATATCCGGTGGTTCCTGTGGCGAGTTCTGCGCGCACGTAAGA 539
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 540 CTGGCGACGCTGTTATCGGTATGGGTGCTGCTGACCGATTCCTGCGTAAAGTTAACCGCATCCGT 599
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 600 TTTAAAGACCATGACTTTTCGCGCTATCGCTGACTTCGACATGGTGGCGTAAGCGCATAGAT 659
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 660 GCACCTAAAGCAGCTGGGTATGATGCTCGCTGGGTAAACCTGTTCTCGCTGACCTGTTT 719
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGlyLysTyrGlyIleLeuGlyValGlu 180
Db 720 TACTCTCGGACGCGGAAATGTTTCGACGTATGATGGAATAATACGCAATTCGCGCGTGGAA 779
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 780 ATGAAGCGGCTGGTATCTACGCGCTCGCTGACAGATTTGGCGGAAAGCCCTGACCATC 839
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 840 TGCACCGTATCTGACCATCTCCGACTCAGCAGCAGACCATCTGCGCTGAGCGTCAGACT 899
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 900 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGGCGATAAAGAG 956
```

RESULT 8

```
US-11-036-497-3
; Sequence 3, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Beetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; TITLE OF INVENTION: and modified analogues thereof
; FILE REFERENCE: 02901/000J410-USO
```

;/ CURRENT APPLICATION NUMBER: US/11/036,497
;/ CURRENT FILING DATE: 2005-01-14
;/ PRIOR APPLICATION NUMBER: PCT/EP99/10416
;/ PRIOR FILING DATE: 1999-12-23
;/ PRIOR APPLICATION NUMBER: M198A002792
;/ PRIOR FILING DATE: 1998-12-23
;/ NUMBER OF SEQ ID NOS: 29
;/ SOFTWARE: PatentIn version 3.1
;/ SEQ ID NO 3
;/ LENGTH: 3383
;/ TYPE: DNA
;/ ORGANISM: Artificial Sequence
;/ FEATURE:
;/ OTHER INFORMATION: plasmid
;/ FEATURE:
;/ NAME/KEY: gene
;/ LOCATION: (231)..(960)
;/ OTHER INFORMATION: deod
;/ US-11-036-497-3

Alignment Scores:
Pred. No.: 3,96e-142 Length: 3383
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 24 Gaps:

US-10-035-300A-2 (1-239) x US-11-036-497-3 (1-3383)

QY 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 240 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 239
QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 300 GGGGACCCGCTGGCTGCGAAGTATATTGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 359
QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrIysGlyArgLysIleSerVal 60
Db 360 AACAGCTTCGGGTATGCTGGGCTTCACCGGTACTTACAAAGGCGCGCAAAATTCGCGTA 419
QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 420 ATGGGTACAGGTATGGGTATCCCGTCTCTGCTCCATCTACACCAAGAACTGATCACCGAT 479
QY 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 480 TTCCGGGTGAAGAAATTAATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCCGCAACGTAAAA 539
QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 540 CTCGGCGACGTGTTATCGGTATGGGTGCTGACCGATTCACCAAGTTTAAACCGATCCCGT 599
QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 600 TTTAAAGACCATGATCTTGGCGGTATCGCTGATCTTCGACATGTTGGGTACGCGTAGAT 659
QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 660 GCAGCTAAAGCACTGGGTATTGATGCTCGGTGGGTAAACCTGTCTCCGCTGACCTGTTTC 719
QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 720 TACTCTCCGACGCGGAAATGTTTCGACGTGATGGAAAAATACGGCATTCCTCGCGGTGAA 779
QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 780 ATGGAAGCGGTGATCTACGCGGTGCTGTCAGAAATTTGGCGCGGAAGCCCTGACCATC 839
QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 840 TGCACCGTATCTGACCACATCCGCACTCACGACGACCACTCCCGCTGAGCGTCAGACT 899

QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 900 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 956

RESULT 9

US-09-891-865A-14
;/ Sequence 14, Application US/09891865A
;/ Publication No. US20030059870A1
;/ GENERAL INFORMATION:
;/ APPLICANT: NORPHARMA SPA
;/ TITLE OF INVENTION: Recombinant bacterial strains for the production of
;/ TITLE OF INVENTION: natural nucleosides and modified analogues thereof
;/ FILE REFERENCE: 99DC26E
;/ CURRENT APPLICATION NUMBER: US/09/891,865A
;/ CURRENT FILING DATE: 2002-06-25
;/ PRIOR APPLICATION NUMBER: M198A002792
;/ PRIOR FILING DATE: 1998-12-23
;/ NUMBER OF SEQ ID NOS: 15
;/ SOFTWARE: PatentIn Ver. 2.1
;/ SEQ ID NO 14
;/ LENGTH: 3934
;/ TYPE: DNA
;/ ORGANISM: Artificial Sequence
;/ FEATURE:
;/ OTHER INFORMATION: Description of Artificial Sequence: udp and deod
;/ OTHER INFORMATION: cloned downstream piac promoter
;/ US-09-891-865A-14

Alignment Scores:

Pred. No.: 4,88e-142 Length: 3934
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps:

US-10-035-300A-2 (1-239) x US-09-891-865A-14 (1-3934)

QY 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 198 ATGGCTACCCACACATTAATGCAGAAATGGCGATTTCGCTGACGTAGTTTGTATGCCA 257
QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 258 GGGGACCCGCTGGCTGCGAAGTATATTGCTGAAACTTTCCTTGAAGATGCCCGTGAAGTG 317
QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 318 AACACAGTTTCGGGTATGCTGGGCTTCACCGGTACTTACAAAGGCGCGCAAAATTCGCGTA 377
QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 378 ATGGGTACAGGTATGGGTATCCCGTCTCTCCATCTACACCAAGAACTGATCACCGAT 437
QY 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 438 TTCCGGGTGAAGAAATTAATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCCGCACTGATAA 497
QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 498 CTCGGCGACGTGTTATCGGTATGGGTGCTGTCACCGATTCACCAAGTTTAAACCGATCCCGT 557
QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 558 TTTAAAGACCATGACTTTGCCGTATCGCTGATCTTCGACATGGTGGGTAAACGCGTAGAT 617
QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 618 GCAGCTAAAGCACTGGGTATTGATGCTCGGTGGGTAACTGTTCTCCGCTGACCTGTTTC 677
QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180

```
Db      678 TACTCTCGGACGGCGAAATGTTCCAGCTGATGGAAAAATACGGCAATTCCTCGGCGTGGAA 737
Qy      181 MetGluAlaAlaGlyIleTyrglyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db      738 ATGGAAGCGGTGGTATCTACGGCGTCTGCGAATTTGGCGGAAAGCCCTGACCATC 797
Qy      201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db      798 TGCACCGTATCTGACCATCGCACTACGAGCAGACCACTCGCGCTGAGCGTCAGACT 857
Qy      221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db      858 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 914

RESULT 10
US-11-036-497-14
; Sequence 14, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppe
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; FILE REFERENCE: and modified analogues thereof
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCR/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 3934
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: udp and deod cloned downstream ptac promoter
US-11-036-497-14

Alignment Scores:
Pred. No.: 4,88e-142 Length: 3934
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 24 Gaps: 0

US-10-035-300A-2 (1-239) x US-11-036-497-14 (1-3934)
Qy      1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db      198 ATGGCTACCCACACATTAATGCGAATAATGGCGGATTCGCTGACGTAGTTTGTATGCGCA 257
Qy      21 GlyAspProLeuArgAlaLysTyriLeAlaGlnThrPheLeuGluAspAlaArgGluVal 40
Db      258 GCGACCGCGCTGCGGCGAAGTATATTCCTGAAACTTTCTTGAAGATGCCCGTGAAGTG 317
Qy      41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyriLysGlyArgLysIleSerVal 60
Db      318 AACAACTTCCGGTATGCTGGCTTCCCGGTACTTACAAGCGCGCAAAATTTCCGTA 377
Qy      61 MetGlyHisGlyValGlyIleProSerCysSerIleTyriThrLysGluLeuIleThrAsp 80
Db      378 ATGGGTCAACGGTATGGGTATCCCGTCTGCTCCATCATCACCAAGAACTGATCACCGAT 437
Qy      81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db      438 TTCGGCGTGAAGAAATTAATCCGGTGGGTTCTGTGGCGGAGTTCTGCCGACGTAAA 497
```

```
Qy      101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db      498 CTGCGCGACGTCGTATTCGGTATCGGTGCTGCACCGATTCCAAAGTTTAAACCGCATCCGT 557
Qy      121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db      558 TTTAAAGACCATGACTTTGCGCGCTATCGCTGACTTCGACATGGTGCCTGAACGCACTAGAT 617
Qy      141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db      618 GCAGCTAAAGCACTGGGTATTGATGCTGCGGTGGGTAACTGTTCTCCGCTGACCTGTTTC 677
Qy      161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyriGlyIleLeuGlyValGlu 180
Db      678 TACTCTCGGACGGCGAAATGTTCCAGCTGATGGAAAAATACGGCAATTCCTCGGCGTGGAA 737
Qy      181 MetGluAlaAlaGlyIleTyrglyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db      738 ATGGAAGCGGTGGTATCTACGGCGTCTGCGAATTTGGCGGAAAGCCCTGACCATC 797
Qy      201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db      798 TGCACCGTATCTGACCATCGCACTACGAGCAGACCACTCGCGCTGAGCGTCAGACT 857
Qy      221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db      858 ACCTTCAACGACATGATCAAAATCGCACTGGAAATCCGTTCTGCTGGCGGATAAAGAG 914

RESULT 11
US-09-891-865A-5
; Sequence 5, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; FILE REFERENCE: natural nucleosides and modified analogues thereof
; CURRENT APPLICATION NUMBER: US/09/891,865A
; CURRENT FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 4189
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; NAME/KEY: gene
; LOCATION: (243)..(1021)
; OTHER INFORMATION: udp
; NAME/KEY: gene
; LOCATION: (1037)..(1766)
; OTHER INFORMATION: deod
US-09-891-865A-5

Alignment Scores:
Pred. No.: 5,32e-142 Length: 4189
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-891-865A-5 (1-4189)
Qy      1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db      1046 ATGGCTACCCACACATTAATGCGAATAATGGCGGATTCGCTGACGTAGTTTGTATGCGCA 1105
```

```
QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
DB 1106 GGGACCCGCTGGTGGCAAGATATATGCTGAACATTTCTTGAAGATGCCCGTGAAGTG 1165

QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
DB 1166 AACAAAGCTTCGGGTATGCTGGGCTTCAACGGTACTTACAAAGCGCCGCAAAATTTCCGTA 1225

QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
DB 1226 ATGGGTACACGGTATGGGTATCCCGTCTGCTGCATCTACACCAAGAACTGATCACCGAT 1285

QY 81 PheGlyValLysIleAlaGlyValGlySerCysGlyAlaValLeuProHisValLys 100
DB 1286 TTGGCGGTGAAGAAATATATCCGGTGGGTCTGTGGCGCAGTTCTGGCGCAGCTAATA 1345

QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
DB 1346 CTGGCGACGCTCGTTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAACCGCATCCGT 1405

QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
DB 1406 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGGCTAAACGAGTAGAT 1465

QY 141 AlaAlaLysAlaLeuGlyIleAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
DB 1466 GCAGCTAAAGCAGCTGGGTATGATGCTGGGTGGGTGCTGCGTGAACCTGCTGCTGTTTC 1525

QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
DB 1526 TACTCTCCGACGCGGAAATTTTCGACGTGATGGAAATACGGCATTTCTCGCGGTGAA 1585

QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
DB 1586 ATGGAAGCGCTGGTATCTACGGCGTCTGCTGCAAGATTTTGGCGGAAAGCCCTGACCATC 1645

QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
DB 1646 TGCACCGTATCTGACCATCTCCGACTTCGACATGGTGGCTGAGCGTCAAGCT 1705

QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
DB 1706 ACCTTCAACGACATGATCAAAATCGCAATCGGATCCGTTCTGCTGGCGGATAAAGAG 1762

RESULT 12
US-11-036-497-5
; Sequence 5, Application US/11036497
; Publication No. US20050142645A1
; GENERAL INFORMATION:
; APPLICANT: Bestetti, Giuseppina
; APPLICANT: Cali', Simona
; APPLICANT: Ghisotti, Daniela
; APPLICANT: Orsini, Gaetano
; APPLICANT: Tonon, Giancarlo
; APPLICANT: Zuffi, Gabriele
; TITLE OF INVENTION: Recombinant bacterial strains for the production of natural nucle
; FILE OF INVENTION: and modified analogues thereof
; FILE REFERENCE: 02901/000J410-US0
; CURRENT APPLICATION NUMBER: US/11/036,497
; CURRENT FILING DATE: 2005-01-14
; PRIOR APPLICATION NUMBER: PCT/EP99/10416
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: MI98A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 4189
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: plasmid
```

```
;
; NAME/KEY: gene
; LOCATION: (243)..(1021)
; OTHER INFORMATION: udp
;
; FEATURE:
;
; NAME/KEY: gene
; LOCATION: (1037)..(1766)
; OTHER INFORMATION: deod
US-11-036-497-5

Alignment Scores: 5,32e-142 Length: 4189
Pred. No.: 1222,00 Matches: 238
Score: 1222,00 Conservative: 1
Percent Similarity: 100,00%
Best Local Similarity: 99,58% Mismatches: 0
Query Match: 99,76% Indels: 0
DB: 24 Gaps: 0

US-10-035-300A-2 (1-239) x US-11-036-497-5 (1-4189)

QY 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
DB 1046 ATGGCTACCCACACATTAATGCAAAATGGCGATTTGCTGACGTAGTTTGTATGCCA 1105

QY 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
DB 1106 GCGCACCCTGGTGGTGAAGTATATTGCTGAACATTTCTTGAAGATGCCCGTGAAGTG 1165

QY 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
DB 1166 AACAAAGCTTCGGGTATGCTGGGCTTCAACGGTACTTACAAAGCGCCGCAAAATTTCCGTA 1225

QY 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
DB 1226 ATGGGTACACGGTATGGGTATCCCGTCTGCTGCATCTACACCAAGAACTGATCACCGAT 1285

QY 81 PheGlyValLysIleAlaGlyValGlySerCysGlyAlaValLeuProHisValLys 100
DB 1286 TTGGCGGTGAAGAAATTTTCGACGTGATGGTGGGTGCTGCGTGAACCTGCTGCGTGAAGT 1345

QY 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
DB 1346 CTGGCGACGCTCGTTATCGGTATGGGTGCTGCACCGATTTCCAAAGTTAACCGCATCCGT 1405

QY 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
DB 1406 TTTAAAGACCATGACTTTGGCGCTATCGCTGACTTCGACATGGTGGCTAAACGAGTAGAT 1465

QY 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
DB 1466 GCAGCTAAAGCAGCTGGGTATGATGCTGGGTGGGTGCTGCGTGAACCTGCTGCTGTTTC 1525

QY 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
DB 1526 TACTCTCCGACGCGGAAATTTTCGACGTGATGGAAATACGGCATTTCTCGCGGTGAA 1585

QY 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
DB 1586 ATGGAAGCGCTGGTATCTACGGCGTCTGCTGCAAGATTTTGGCGGAAAGCCCTGACCATC 1645

QY 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
DB 1646 TGCACCGTATCTGACCATCTCCGACTTCGACATGGTGGCTGAGCGTCAAGCT 1705

QY 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
DB 1706 ACCTTCAACGACATGATCAAAATCGCAATCGGATCCGTTCTGCTGGCGGATAAAGAG 1762

RESULT 13
US-10-198-034-5
; Sequence 5, Application US/10198034
; Publication No. US20030077268A1
; GENERAL INFORMATION:
```


Qy 121 PheLysAspHisAspPheAlaAlaAlaAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 773 TTTTAAAGACCAATGACTTTGGCGCTATCGCTGACTTCGACATGGTGCGTAACGCAATAGAT 832
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 833 GCAGCTAAAGCACCTGGGTATTGATGCTCGGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 892
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 893 TACTCTCCGACCGCGCAATGTTTCGACGTGATGGAATAATACGCATTCCTCGGCGTGGAA 952
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 953 ATGGAACGGCTGTGATCTACGGCGTTCGCTGCAAAATTTGGCGGAAGCCCTGACCATC 1012
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1013 TGCACCGTATCTGACCATCCGACCTCAGCAGCAGACCACTGCGCTGAGCGTCCAGACT 1072
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1073 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 1129

RESULT 15

US-09-891-865A-7
; Sequence 7, Application US/09891865A
; Publication No. US20030059870A1
; GENERAL INFORMATION:
; APPLICANT: NORPHARMA SPA
; TITLE OF INVENTION: Recombinant bacterial strains for the production of
; TITLE OF INVENTION: natural nucleosides and modified analogues thereof
; FILE REFERENCE: 99DC26E
; CURRENT APPLICATION NUMBER: US/09/891.865A
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: M198A002792
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 5241
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Plasmid
; FEATURE:
; NAME/KEY: gene
; LOCATION: (1312)..(2042)
; OTHER INFORMATION: deod
US-09-891-865A-7

Alignment Scores:
Pred. No.: 7.26e-142 Length: 5241
Score: 1222.00 Matches: 238
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.58% Mismatches: 0
Query Match: 99.76% Indels: 0
DB: 10 Gaps: 0

US-10-035-300A-2 (1-239) x US-09-891-865A-7 (1-5241)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 1322 ATGGCTACCCACACATTAATGCAAGAAATGGCGAATTCGCTGACGTAGTITTTGATGCCA 1381
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 1382 GGCAGCCCGCTGCGTGGGAAGTATATTGCTGAACCTTTCTTGAAGATGCCCGTGAAGTG 1441
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 1442 AACAAACGTTCCGCGTATGCTGGGCTTCACCGGTACTTACAAAGCGCGCAAAATTTCCGTA 1501

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 1502 ATGGGTACAGGTATGGGTATCCCGTCTCTGCTCATCTACACCAAGACTGATCACCAGAT 1561
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 1562 TTCGGCGTGAAGAAAATTTATCCGCGTGGGTTCCTGTGGCGCAGTTCTGCGCGCACGTAA 1621
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 1622 CTGCGCGACGTCTGTTATCGGTATGGGTGCTGTCACCGATTTCCAAAGTTAACCGCATCCGT 1681
Qy 121 PheLysAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 1682 TTTTAAAGACCATGACTTTTGGCGCTATCGCTGACTTCGACATGGTGGTAAACGCAATAGAT 1741
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 1742 GCAGCTAAAGCACCTGGGTATTGATGCTCGCGTGGGTAAACCTGTTCTCCGCTGACCTGTTTC 1801
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 1802 TACTCTCCGACCGCGAAAATGTTTCGACGTGATGGAATAATACGCATTTCTCGCGGTGGAA 1861
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 1862 ATGGAAGCGCTGCTGATCTACGGCGTTCGCTGCAAGAAATTTGGCGGAAAGCCCTGACCATC 1921
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 1922 TGCACCGTATCTGACCACTCCGACCTCAGCAGCAGACCACTGCGCGTGAAGCGTCAGACT 1981
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeuLeuGlyAspLysGlu 239
Db 1982 ACCTTCAACGACATGATCAAAATCGCACTGGAATCCGTTCTGCTGGCGGATAAAGAG 2038

Search completed: July 23, 2005, 06:11:47

Job time : 757 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:23:45 ; Search time 39 Seconds
(without alignments)
589.636 Million cell updates/sec

Title: US-10-035-300A-2
Perfect score: 1225
Sequence: 1 MATPHINAEMGDFADVVLMP.....TTFNDMIKIALESVLLGDKE 239
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1222	99.8	239	2	G91296
2	1222	99.8	239	2	A27854
3	1222	99.8	239	2	A86138
4	1187	96.9	239	2	AF1073
5	1106	90.3	239	2	AF0054
6	1085	88.6	239	2	JC4326
7	1017	83.0	238	2	B64074
8	993	81.1	241	2	C82087
9	858	70.0	245	2	C82505
10	778	63.5	234	2	A84993
11	693.5	56.6	233	2	D59614
12	675	55.1	233	2	B64667
13	675	55.1	305	2	D75306
14	671.5	54.8	233	2	AH1678
15	670.5	54.7	233	2	AH1306
16	663	54.1	233	2	C71848
17	659.5	53.8	234	2	JT0874
18	659	53.8	234	2	C86741
19	576.5	47.1	236	2	B97964
20	575.5	47.0	236	2	E95096
21	569	46.4	236	2	G30007
22	565.5	46.2	235	2	D89774
23	555.5	45.3	233	2	C71288
24	468	38.2	320	2	D64205
25	450	36.7	238	2	S73418
26	430.5	35.1	239	2	G82909
27	389	31.8	232	2	E30592
28	288.5	23.6	236	2	S58291
29	288	23.5	273	2	A84246

30 282 23.0 258 2 D84336 uridine phosphoryl
31 270 22.0 254 2 D72516 probable uridine p
32 246.5 20.1 238 2 G86729 uridine phosphoryl
33 244.5 20.0 253 2 G86070 uridine phosphoryl
34 244.5 20.0 253 2 A98224 uridine phosphoryl
35 243.5 19.9 253 2 AF0916 uridine phosphoryl
36 243.5 19.9 253 2 T46830 uridine phosphoryl
37 243.5 19.9 253 2 S05491 uridine phosphoryl
38 241.5 19.7 253 2 A10460 uridine phosphoryl
39 238 19.4 263 2 A72697 probable purine nu
40 237.5 19.4 258 2 B82249 uridine phosphoryl
41 234 19.1 247 2 AC0923 probable uridine p
42 231 18.9 259 2 C64059 uridine phosphoryl
43 202 16.5 258 2 F71251 probable uridine p
44 175.5 14.3 243 2 F82497 uridine phosphoryl
45 175.5 14.3 279 2 JC4327 uridine phosphoryl

RESULT 1
G91296
Purine-nucleoside phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: G91296
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G91296
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <HAY>
A;Cross-references: UNIPROT:P09743; GB:BA000007; PIDN:BA838766.1; PID:gl3364821; GSPDB:G A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs5343
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 99.8%; Score 1222; DB 2; Length 239;
Best Local Similarity 99.6%; Pred. No. 4.7e-95;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVVLMPGDPPLAKYIAETFLIEDAREVNNVRGMLGFTCTYKGRKISV 60
Db 1 MATPHINAEMGDFADVVLMPGDPPLAKYIAETFLIEDAREVNNVRGMLGFTCTYKGRKISV 60
Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIIVGMGACTDSKVNRI 120
Db 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIIVGMGACTDSKVNRI 120
Qy 121 FKDHFAAIDAFDMVRNVAADKALGIDARVGNLFSADLFYSPDGEPMFVNMKEYGILGVE 180
Db 121 FKDHFAAIDAFDMVRNVAADKALGIDARVGNLFSADLFYSPDGEPMFVNMKEYGILGVE 180
Qy 181 MEANGIYGVAAEFGAKALTCTVSDHRTHEOTTAARQTTFNDMIKIALESVLLGDKE 239
Db 181 MEANGIYGVAAEFGAKALTCTVSDHRTHEOTTAARQTTFNDMIKIALESVLLGDKE 239

RESULT 2
A27854
Purine-nucleoside phosphorylase (EC 2.4.2.1) - Escherichia coli (strain K-12)
N;Alternate names: inosine phosphorylase
C;Species: Escherichia coli
C;Date: 31-Mar-1988 #sequence_revision 03-Aug-1992 #text_change 09-Jul-2004
C;Accession: A41143; A27854; S56608; G65253
R;Hershfield, M.S.; Chaffee, S.; Koro-Johnson, L.; Mary, A.; Smith, A.A.; Short, S.A. Proc. Natl. Acad. Sci. U.S.A. 88, 7185-7189, 1991
A;Title: Use of site-directed mutagenesis to enhance the epitope-shielding effect of cov

A;Reference number: A41143; MUID:91334430; PMID:1714590
A;Accession: A41143
A;Molecule type: DNA
A;Residues: 1-239 <HER>
A;Cross-references: UNIPROT:P09743; GB:M60917; NID:g147308; PIDN:AAA24401.1; PID:g147309
R;Larsen, J.E.L.; Albrechtsen, B.; Valentin-Hansen, P.
Nucleic Acids Res. 15, 5125-5140, 1987
A;Title: Analysis of the terminator region after the deoCABD operon of Escherichia coli
A;Reference number: A27854; MUID:87259990; PMID:3299264
A;Accession: A27854
A;Molecule type: DNA
A;Residues: 225-239 <LAR>
A;Cross-references: GB:X05629; NID:g41254; PIDN:CAA29114.1; PID:g41255
R;Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A;Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A;Reference number: S56314; MUID:95334362; PMID:7610040
A;Accession: S56608
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-239 <BUR>
A;Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97280.1; PID:g537224
A;Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1994
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Ferna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G65253
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-239 <BLAT>
A;Cross-references: GB:AE000508; GB:U00096; NID:g2367382; PIDN:NAC77337.1; PID:g1790844;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: deoD
A;Map position: 100
A;Superfamily: purine-nucleoside phosphorylase pnp
C;Keywords: glycosyltransferase; pentosyltransferase; salvage pathway
Query Match 99.8%; Score 1222; DB 2; Length 239;
Best Local Similarity 99.6%; Pred. No. 4.7e-95;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRCMLGFTGTYGKRKISV 60
Db 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRCMLGFTGTYGKRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVIWGMGACTDSKVNRI 120
Db 61 MGHGMGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVIWGMGACTDSKVNRI 120
QY 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYIGLVE 180
Db 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYIGLVE 180
QY 181 MEAAGYGVAAEFGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
Db 181 MEAAGYGVAAEFGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
RESULT 3
A86138
purine-nucleoside phosphorylase [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: A86138
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A86138

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <STO>
A;Cross-references: UNIPROT:P09743; GB:AE005174; NID:g12519411; PIDN:AAG59565.1; GSPDB:GN
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: deoD
A;Superfamily: purine-nucleoside phosphorylase pnp
Query Match 99.8%; Score 1222; DB 2; Length 239;
Best Local Similarity 99.6%; Pred. No. 4.7e-95;
Matches 238; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRCMLGFTGTYGKRKISV 60
Db 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRCMLGFTGTYGKRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVIWGMGACTDSKVNRI 120
Db 61 MGHGMGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVIWGMGACTDSKVNRI 120
QY 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYIGLVE 180
Db 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYIGLVE 180
QY 181 MEAAGYGVAAEFGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
Db 181 MEAAGYGVAAEFGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
RESULT 4
AF1073
purine nucleoside phosphorylase [EC 2.4.2.1] [imported] - Salmonella enterica subsp. ent
C;Species: Salmonella enterica subsp. enterica serovar typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AF1073
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
.S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AF1073
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-239 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD03405.1; PID:g16505674; GSPDB:GN00176
C;Genetics:
A;Gene: deoD
A;Superfamily: purine-nucleoside phosphorylase pnp
C;Keywords: glycosyltransferase; pentosyltransferase
Query Match 96.9%; Score 1187; DB 2; Length 239;
Best Local Similarity 96.7%; Pred. No. 4e-92;
Matches 231; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRCMLGFTGTYGKRKISV 60
Db 1 MATPHINAEMGDFADVVLMPGDPPLRAKYIAETPLEDAREVNNVRCMLGFTGTYGKRKISV 60
QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVIWGMGACTDSKVNRI 120
Db 61 MGHGMGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVIWGMGACTDSKVNRI 120
QY 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYIGLVE 180
Db 121 FKDHDFAAIADFDMVRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYIGLVE 180
QY 181 MEAAGYGVAAEFGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239
Db 181 MEAAGYGVAAEFGAKALITCTVSDHIRTHTHEQTAAEROTTENDMIKIALESVLLGDKE 239

Query Match	88.6%	Score 1085;	DB 2;	Length 239;
Best Local Similarity	90.4%	Pred. No. 1.4e-83;		
Matches 217;	Conservative	7;	Mismatches 14;	Indels 2;
				Gaps 2

RESULT 8
C82087
purine nucleoside phosphorylase VC2347 [imported] - Vibrio cholerae (strain N1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: C82087
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson
charadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.;
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio chol
A;Reference number: A82015; MUID:20406833; PMID:10952301

A;Accession: C82087
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-241 <HEI>
A;Cross-references: UNIPROT:Q9KPM0; GB:AE004305; GB:AE003852; NID:g9656912; PIDN:AAF9549
A;Experimental source: serogroup O1; strain N16961; biotype E1 Tor
C;Genetics:
A;Gene: VC2347
A;Map position: 1
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 81.1%; Score 993; DB 2; Length 241;
Best Local Similarity 79.8%; Pred. No. 7.2e-76;
Matches 190; Conservative 26; Mismatches 22; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDPADVLMFGDPLRAKYIAETFLDAREVNNVRLMGFTGTGKGRKISV 60
Db 1 MATPHINAQMGDPADVLMFGDPLRAKYIAENFLDNAVQCDVRNMFVGTGTGKGRKISV 60

QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 61 MGHGMGIPSCSIYTKELIDYGVKKIIRVSCGAVNEGIVKRVLDVLVGMGACTDSKNRIR 120

QY 121 FKDHFAALADFDWVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVMEYKILGVE 180
Db 121 FKDHFAALADYKMWKAAEEAAKARGIDVKVGNLFSALFYTPDPSMFDMVKYIGVGE 180

QY 181 MEAAGIYGVAAEFGAKALITCTVSDHIRTQTTAAERQTTFNMDMKIALESVLLGDK 238
Db 181 MEAAGIYGVAAYGAKALAITCTVSDHIKTGEQTTSEERQNTFNEMIEIALDSVLIGDQ 238

RESULT 9
C82505
purine nucleoside phosphorylase VCA0053 [imported] - Vibrio cholerae (strain N16961 sero
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
A;Accession: C82505
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bae, S.; Qin, H.; Dragoi, I.; Sellers, F.
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: C82505
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-245 <HEI>
A;Cross-references: UNIPROT:Q9KNB2; GB:AE004349; GB:AE003853; NID:g9657434; PIDN:AAF9596
A;Experimental source: serogroup O1; strain N16961; biotype E1 Tor
C;Genetics:
A;Gene: VCA0053
A;Map position: 2
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 70.0%; Score 858; DB 2; Length 245;
Best Local Similarity 69.2%; Pred. No. 1.5e-64;
Matches 182; Conservative 36; Mismatches 36; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDPADVLMFGDPLRAKYIAETFLDAREVNNVRLMGFTGTGKGRKISV 60
Db 10 MATPHINAQMGDPADVLMFGDPLRAKYIAETFLDVKQCDVRSMFGFTGTGKGRKISV 69

QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 70 MGHGMGIPSCSIYTKELIAYGVKNIRIGSCGAVRDDVKLMDVWVGMGASTDSKNRIR 129

QY 121 FKDHFAALADFDWVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVMEYKILGVE 180
Db 130 FSGHFAALADYDLLETAVNQAQQVQVKNVGFSDLFYTPPEPIFEKMKLGLIGVD 189

QY 181 MEAAGIYGVAAEFGAKALITCTVSDHIRTQTTAAERQTTFNMDMKIALESVLL 234
Db 181 MEAAGIYGVAAYGAKALAITCTVSDHIKTGEQTTSEERQNTFNEMIEIALDSVLIGDQ 238

Db 190 MEAAGIYGVAADIGARALTILTVSDHILRGKLSSEDSROKSFNDMMKVALETAI 243

RESULT 10
A84993
purine-nucleoside phosphorylase (EC 2.4.2.1) [imported] - Buchnera sp. (strain APS)
C;Species: Buchnera sp.
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: A84993
R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A;Reference number: A84930; MUID:20445173; PMID:10993077
A;Accession: A84993
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-234 <STO>
A;Cross-references: GB:AP000398; GSPDB:GN00144
A;Experimental source: strain APS
C;Genetics:
A;Gene: deoD; BU541
C;Superfamily: purine-nucleoside phosphorylase pnp
C;Keywords: glycosyltransferase; pentosyltransferase

Query Match 63.5%; Score 778; DB 2; Length 234;
Best Local Similarity 61.5%; Pred. No. 7.1e-58;
Matches 144; Conservative 43; Mismatches 47; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDPADVLMFGDPLRAKYIAETFLDAREVNNVRLMGFTGTGKGRKISV 60
Db 1 MSTPHINSKDDFSDIVLMFGDPLRAKYIAEYKLSNFVQVNDTRLVAVTFYKNRKISI 60

QY 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 61 MSHGIGIPASLYTRELIIEFNFKIIRITGTCGAVRDDIKLRDIVSMGASTDSKNRIR 120

QY 121 FKDHFAALADFDWVRNAVDAKALGIDARVGNLFSADLFYSPDGMFDMVMEYKILGVE 180
Db 121 FNDHFAALADFDWVYIVVISKMKIKVIGNFFTTDSFYNDKQMLILKYNILGVD 180

QY 181 MEAAGIYGVAAEFGAKALITCTVSDHIRTQTTAAERQTTFNMDMKIALESVLL 234
Db 181 METAGIYGVAELKQALSICTVSDHITNKFLSKERESSFNDMIELALESVLL 234

RESULT 11
D69614
purine nucleoside phosphorylase deoD - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: D69614
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallerc
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.;
Koster, P.; Konningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, P.; Pohl, T.M.; Portetelle,
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror,
teuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpsira, P.; Tognoni, A.; Torato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: D69614
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <KUN>
A;Cross-references: UNIPROT:O34925; GB:Z99114; GB:AL009126; NID:g2634230; PIDN:CAB13854-J
A;Experimental source: strain 168

C;Genetics: A;Gene: deod C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 56.6%; Score 693.5; DB 2; Length 233;
Best Local Similarity 58.1%; Pred. No. 8.5e-51; Mismatches 54; Indels 1; Gaps 1;
Matches 133; Conservative 41

Qy 5 HINAEAGDPADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMHG 64
Db 4 HIGAEKGQIADTVLLPGDPLRAKIAETFLDAREVNNVRGMLGFTGTGKGRKISVMHG 63

Qy 65 VGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGVMGACTDTSKVNIRPKDH 124
Db 64 MGVPISIIYVNELIQSDVQNLIRVSCGAVLPHVKLRDVGVMGACTDTSKVNIRPKDH 123

Qy 125 DFAIAIDPDMVRNADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVMEAA 184
Db 124 DFACADPELUNAYDAKQGVPTVGSVFADQFTDNDSDQI-EKLUKYGVLGVMEETT 182

Qy 185 GIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESV 233
Db 183 ALYTLAAKHGRKALSILTVSDHVLTGERTTAERQTTFNMDIKIALESV 231

RESULT 12
B64667
purine-nucleoside phosphorylase - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C;Accession: B64667
R;Tomb, J.F.; White, O.; Kervage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C. A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: B64667
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <TOM>
A;Cross-references: UNIPROT:P56463; GB:AE000623; GB:AE000511; NID:g2314327; PIDN:AA0822
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 55.1%; Score 675; DB 2; Length 233;
Best Local Similarity 55.0%; Pred. No. 3e-49; Mismatches 61; Indels 0; Gaps 0;
Matches 126; Conservative 42

Qy 3 TPHINAEAGDPADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMG 62
Db 2 TPHINAKIGDFYPQCLCGDPLRVSYIAKFLQDAKEITNVNRNMLGFGSKYKGRGISLMG 61

Qy 63 HGVIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGVMGACTDTSKVNIRPK 122
Db 62 HGMGIASCTIYVTELIQYQKELIRIGTCGAISPKVGLGDIIMATGASTDTSKVNIRPK 121

Qy 123 DHDFAAIADPDMVRNADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVME 182
Db 122 NHDLSATPDPFELSRAVQTAKRLGIDLKGVNFSDFYFSTHAFDLMAKNHLAEME 181

Qy 183 AAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESV 231
Db 182 AAGLYATAMELNAKALCLCSVDHLITKEALSPKRVESPDNMIILALE 230

RESULT 13
D75306
purine nucleoside phosphorylase - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: D75306
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: D75306
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-305 <WHI>
A;Cross-references: UNIPROT:Q9RSF8; GB:AE002050; GB:AE000513; NID:g6459965; PIDN:AAF11716
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2166
A;Map position: 1
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 55.1%; Score 675; DB 2; Length 305;
Best Local Similarity 54.8%; Pred. No. 4.1e-49; Mismatches 43; Indels 0; Gaps 0;
Matches 126; Conservative 43

Qy 5 HINAEAGDPADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMHG 64
Db 70 HINAEKGQIADTVLLPGDPLRAKIAETFLDAREVNNVRGMLGFTGTGKGRKISVMHG 129

Qy 65 VGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGVMGACTDTSKVNIRPKDH 124
Db 130 MGIPSCMIYVNELIQDYCKNLIRVGTAGSYQEGVHVRDLVLAQAQACTSDSNINRVFGR 189

Qy 125 DFAIAIDPDMVRNADAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVMEAA 184
Db 190 TFPIADPELUNAYDAKQGVPTVGSVFADQFTDNDSDQI-EKLUKYGVLGVMEETT 249

Qy 185 GIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESV 234
Db 250 GLYTLAAKYGVRALITLITSDHVLTHVETSAREERQLTFNGMIEVALDAAL 299

RESULT 14
AH1678
purine nucleoside phosphorylase [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AH1678
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.; D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehland, A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1678
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-233 <GLA>
A;Cross-references: UNIPROT:Q92AF2; GB:AL592022; PIDN:CAC97200.1; PID:gl6414471; GSPDB:G3
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: deod
C;Superfamily: purine-nucleoside phosphorylase pnp

Query Match 54.8%; Score 671.5; DB 2; Length 233;
Best Local Similarity 56.6%; Pred. No. 5.9e-49; Mismatches 133; Conservative 36; Indels 11; Gaps 2;
Matches 133; Conservative 36

Qy 5 HINAEAGDPADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGTGKGRKISVMHG 64
Db 4 HIEAKQGEIAETILLPGDPLRAKIAETFLDAREVNNVRGMLGFTGTGKGRKISVMHG 63

Qy 65 VGIPSCSIYTKELITDFGVKKIIRVSCGAVLPHVKLRDVGVMGACTDTSKVNIRPKDH 124
Db 64 MGIPSIYVNELIQSDVQNLIRVGTAGSYQEGVHVRDLVLAQAQACTSDSNINRVFGR 123

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 22, 2005, 23:48:48 ; Search time 3877 Seconds
(without alignments)
2346.495 Million cell updates/sec

Title: US-10-035-300A-2

Perfect score: 1225
Sequence: 1 MATPHINAEMGDFADVVLMP.....TTFNDMIKIALESVLLGDKE 239

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool_p/US10035300/runat_18072005_165351_27984/app_query.fasta_1.391
-DB=EST -QFMT=fastcap -SURFIX=rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USERS=US10035300 @CGN 1 1 5180 @runat_18072005_165351_27984 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_hcc.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	683	55.8	771	7	CV219418 EST879128
C 2	683	55.8	794	7	CV219419 EST879129
C 3	681	55.6	767	7	CV205284 EST864994
C 4	681	55.6	769	7	CV204676 EST864386
C 5	681	55.6	792	7	CV204675 EST864385
C 6	681	55.6	794	7	CV205285 EST864995
C 7	668	54.5	721	7	C0577885 TVEST085E
C 8	591	48.2	1589	8	BH770690 LLMGtag44
C 9	436.5	35.6	878	8	AZ533877 ENTBJ777R

10	436	35.6	921	8	BH138974	BH138974	ENTOV94TF
C 11	429.5	35.1	863	8	AZ680406	AZ680406	ENTGM40TF
C 12	376	30.7	861	8	BH166036	BH166036	ENTSV60TF
C 13	364	29.7	864	8	AZ675191	AZ675191	ENTFV59TR
C 14	343.5	28.0	865	8	AZ669801	AZ669801	ENTLL39TF
C 15	314.5	25.7	882	8	AZ681446	AZ681446	ENTMK91TR
C 16	296	24.2	879	8	AZ533664	AZ533664	ENTFT92TF
C 17	285	23.3	169	7	CF306721	CF306721	HDAL--04-
C 18	285	23.3	169	7	CF306860	CF306860	HDAL--05-
C 19	260	21.2	904	8	AZ536366	AZ536366	ENTCK45TF
C 20	242	19.8	161	7	CF307418	CF307418	HDAL--06-
C 21	239	19.5	861	8	AZ669012	AZ669012	ENTKL76TF
C 22	229.5	18.7	852	8	AZ542090	AZ542090	ENTDY06TF
C 23	223	18.2	801	9	CC924665	CC924665	t078f14da
C 24	212	17.3	626	8	BH023711	BH023711	U1-revers
C 25	212	17.3	637	8	BH023716	BH023716	U4-revers
C 26	207.5	16.9	1189	9	CG754454	CG754454	P049-4-E1
C 27	205	16.7	898	8	BH148646	BH148646	ENTPE80TR
C 28	198	16.2	690	4	BH162943	BH162943	EST565466
C 29	197	16.1	734	4	BH159186	BH159186	EST561709
C 30	196	16.0	709	4	BH160915	BH160915	EST563438
C 31	196	16.0	742	4	BH170790	BH170790	EST573313
C 32	195	15.9	635	2	BF297837	BF297837	056PBH03
C 33	195	15.9	645	4	BH170428	BH170428	EST572951
C 34	195	15.9	768	4	BH159582	BH159582	EST562105
C 35	191.5	15.6	715	9	AG266877	AG266877	Finegold1
C 36	189	15.4	740	8	BH023715	BH023715	U4-forward
C 37	186	15.2	648	2	BF298132	BF298132	060PBG02
C 38	185.5	15.1	737	8	BH023713	BH023713	U3-forward
C 39	185	15.1	604	4	BH162201	BH162201	EST564724
C 40	182.5	14.9	890	6	CD344467	CD344467	ETCVC88TF
C 41	178	14.5	549	6	CD344467	CD344467	ETCVC88TF
C 42	178	14.5	557	6	CD343688	CD343688	ETCVC88TF
C 43	177	14.4	226	7	CV213810	CV213810	EST873520
C 44	177	14.4	250	7	CV213811	CV213811	EST873521
C 45	177	14.4	572	4	BH165171	BH165171	EST567694

ALIGNMENTS

RESULT 1
CV219418/c
LOCUS CV219418 771 bp mRNA linear EST 16-SEP-2004
DEFINITION EST879128 non-normalized T1 cDNA library Trichomonas vaginalis CDNA
clone TVTCX26 3' end, mRNA sequence.
ACCESSION CV219418 GI:52166398
VERSION CV219418.1
KEYWORDS EST.
SOURCE Trichomonas vaginalis
ORGANISM Trichomonas vaginalis
Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
Trichomonadidae; Trichomonadinae; Trichomonas.
REFERENCE 1 (bases 1 to 771)
AUTHORS Carlton,J.M., Dyllal,S., Johnson,P.J. and Fraser,C.M.
TITLE The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis
JOURNAL Unpublished (2004)
COMMENT Other ESTs: EST879129
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seq primer: polydT 24 base.
Location/Qualifiers
1. 771
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TVTCX26"

/clone lib="non-normalized T1 cDNA library"
/note=Tvector: Lambda Triplex2; Site 1: SfiA; Site 2:
SfiB; T. vaginalis strain T1 library constructed from
cDNA, made in lambda Triplex2. Inserts cloned
unidirectionally in the SfiA and SfiB sites. Mass excision
of library produced inserts in pTriplex2 plasmid. Inserts
sequenced from both 5' and 3' ends using Triplex2
sequencing primer and polydT 24 bp primer respectively."

ORIGIN

Alignment Scores:
Pred. No.: 1.77e-75 Length: 771
Score: 683.00 Matches: 134
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 57.26% Mismatches: 58
Query Match: 55.76% Indels: 0
DB: 7 Gaps: 0

US-10-035-300A-2 (1-239) x CV219418 (1-771)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 759 ATGCAACACCCCACTAATCTGCAAGGTTGGCGATTTCGTGAGACAGTTCATGTGC 700

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 699 GCGATCCATCTCGTGCAGACTCATTTGCTGTGTAACACTATCTTGAGACGCTAAGCAATC 640

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 639 AACTCTGCTGTAATGCTTGGTTTACAGGCACATACAGGCGAAACCACTCTCTGTT 580

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 579 ATGGGTATGCGATGGGTATCCCATCTTCCATCTACGCTGAAGAACTCTCAACAGTT 520

Qy 81 PheGlyValLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 519 TACAAGGTTAAGACCATCATCTGTCGACACTTGGCGACAGTTGATCCAAATGTTAC 460

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 459 GTCCGCGATGCTGCAATGTGCAGCATCTGGCAGCATACAAATGTTAAACAGATCGC 400

Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 399 CTTCCTGGCCACGATTTCCGAGCTCAGCCCACTTCGAAGTCGTTAGTCCTTAGTTGAA 340

Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 339 TCCGCCAAGGCTCTCAACATCCCAACACAGGTTGGAAAGGCCTACTCAACAGATATCTC 280

Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 279 TACAGCAAGGAACAAAGGATTAATAGAGCGCTTGTCTCAGTACCACCTTCATTTGCTGCGAA 220

Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 219 ATGGAATCTGCTGCTCCCTTCCCAATTCGTGATAACTATATGCGCAAGAGTCGCTGATC 160

Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 159 TGCACAGTTTCGGATCATCATCACACAGAGTCCGCTACACAGAGAGCGCCAGACA 100

Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeu 234
Db 99 TCATTCCAGAACATGATCAAGATCGCTCTTTGAAGCTACATTG 58

RESULT 2
CV219419
LOCUS CV219419 794 bp mRNA linear EST 16-SEP-2004
DEFINITION EST879129 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
clone TVT26 5' end, mRNA sequence.

CV219419
CV219419.1 GI:52166399
EST.
KEYWORDS Trichomonas vaginalis
SOURCE Trichomonas vaginalis
ORGANISM Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
Trichomonadidae; Trichomonadinae; Trichomonas.
REFERENCE 1 (bases 1 to 794)
AUTHORS Carlon,J.M., Dyall,S., Johnson,P.J. and Fraser,C.M.
TITLE The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis
JOURNAL Unpublished (2004)
COMMENT Other ESTs: EST879128
Contact: Jane Carlon
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlon@tigr.org
Seq primer: lambda Triplex2.
Location/Qualifiers
1..794
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TVT26"
/clone_lib="non-normalized T1 cDNA library"
/notes="Vector: Lambda Triplex2; Site 1: SfiA; Site 2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiA and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

ORIGIN
Alignment Scores:
Pred. No.: 1.85e-75 Length: 794
Score: 683.00 Matches: 134
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 57.26% Mismatches: 58
Query Match: 55.76% Indels: 0
DB: 7 Gaps: 0

US-10-035-300A-2 (1-239) x CV219419 (1-794)

Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 13 ATGCAACACCCCACTAATCTGCAAGGTTGGCGATTTCGTGAGACAGTTCATGTGC 72

Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 73 GCGATCCATCTCGTGCAGACTCATTTGCTGTGTAACACTATCTTGAGAACGCTAAGCAATC 132

Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 133 AACTCTGCTGTAATGCTTGGTTTACAGGCACATACAGGCGAAACCACTCTCTGTT 192

Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 193 ATGGGTATGCGATGGGTATCCCATCTTCCATCTACGCTGAAGAGCTTCAACAGTT 252

Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 253 TACAAGGTTAAGACCATCATCTCGTGTGCGCACTTGGCGCACAGTTGATCCAAATGTTTCA 312

Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 313 GTCCGCGATGCTGCAATGTGCAGCATCTGGCAGCATCTCAATGTTTACAGAAATGCGC 372

Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140

Db	95	TCATTCAGAACCAATGATCAAGATCGCTTCTTGAGCTACATG	54
RESULT 4			
CV204676/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seq primer: lambda Triplex2.

FEATURES

source
Location/Qualifiers
1..769
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TVTA905"

/note="non-normalized T1 cDNA library"
SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiA and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."

ORIGIN

Alignment Scores:
Pred. No.: 3.16e-75 Length: 769
Score: 681.00 Matches: 134
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 57.26% Mismatches: 58
Query Match: 55.59% Indels: 0
DB: 7 Gaps: 0

US-10-035-300A-2 (1-239) x CV204676 (1-769)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 757 ATGGCAACACCCCACTACTGCGAAGGTGGCGATTTCGCTGAGACAGTCTCATGTGC 698
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 697 GGGCATCACTCCGTCGCAAGCTCATTCGTGATAAATCATCTTGAGAACGCTAAGCAAGTC 638
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 637 AACTCTGTCGTGAATGCTAGGTGTTTACAGGCACATACAGGCGCAACCACTCTCTGTT 578
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 577 ATGGGTATGGCATGGGTATCCCATCCATTCATCTACGCTGAGAACTCTACAAGTT 518
Qy 81 PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 517 TACAAGGTTAAGACCATCATCCGTGCGCAGTTCGCGCACAGTTGATCCAAATGTTTAC 458
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 457 GTCGCGCATGTCGTCATGTGCACAGCATCTGGCACAGATTCAAATGTTAACAAGATCGC 398
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 397 CTTCCTGGCCACGATTTCCAGCTTACAGCCCACTTCGAAGTCGTAGTCCTTAGTTGAA 338
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 337 TCGGCCAAGGCTCTCAACATCCCAACACAGGTTGGAAGGCGCTACTCAACAGATATCTTC 278
Qy 161 TyrSerProAspGlyGluMetPheAspValMetClnLysTyrGlyIleLeuGlyValGlu 180
Db 277 TACGAAGGAACAAGGATTAATAGGCCCTTGCTCAGTACCACTTCATCTGCTGCGAA 218
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 217 ATGGATCTGCTGCTCCCTTCCCAATGCTGATTACTATGCGCAAGAGCTGGCTGCATC 158
Qy 201 CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
```

```
Db 157 TGCACAGTTCGGATCACAATCATCACACAGAGTCGCTACACGGGAAGAGCGCCAGACA 98
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeu 234
Db 97 TCATTCAGAACATGATCAAGATCGCTCTCTTGAAGCTACATTG 56

RESULT 5
CV204675 792 bp mRNA linear EST 16-SEP-2004
LOCUS EST864385 non-normalized T1 cDNA library Trichomonas vaginalis CDNA
DEFINITION clone TVTA905 3' end, mRNA sequence.
ACCESSION CV204675
VERSION 1 GI:52151655
KEYWORDS Trichomonas vaginalis
SOURCE Trichomonas vaginalis
ORGANISM Trichomonas vaginalis
REFERENCE 1 (bases 1 to 792)
AUTHORS Carlton, J.M., Dyall, S., Johnson, P.J. and Fraser, C.M.
TITLE The complete genome sequence of the sexually transmitted parasite Trichomonas vaginalis
JOURNAL Unpublished (2004)
COMMENT Other ESTs: EST864386
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seq primer: polydT 24 base.
FEATURES
source
1..792
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone="TVTA905"
/clone_lib="non-normalized T1 cDNA library"
/note="Vector: Lambda Triplex2; Site_1: SfiA; Site_2: SfiB; T. vaginalis strain T1 library constructed from cDNA, made in lambda Triplex2. Inserts cloned unidirectionally in the SfiA and SfiB sites. Mass excision of library produced inserts in pTriplex2 plasmid. Inserts sequenced from both 5' and 3' ends using Triplex2 sequencing primer and polydT 24 bp primer respectively."
```

ORIGIN

Alignment Scores:
Pred. No.: 3.29e-75 Length: 792
Score: 681.00 Matches: 134
Percent Similarity: 75.21% Conservative: 42
Best Local Similarity: 57.26% Mismatches: 58
Query Match: 55.59% Indels: 0
DB: 7 Gaps: 0

US-10-035-300A-2 (1-239) x CV204675 (1-792)

```
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 13 ATGGCAACACCCCACTACTGCGAAGGTGGCGATTTCGCTGAGACAGTCTCATGTGC 72
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 73 GGGCATCACTCCGTCGCAAGCTCATTCGTGATAAATCATCTTGAGAACGCTAAGCAAGTC 132
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 133 AACTCTGTCGTGAATGCTAGGTGTTTACAGGCACATACAGGCGCAACCACTCTCTGTT 192
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
```

```
Db 193 ATGGGTATGGCATGGGTATCCCATCCATTTCCATCTACGCTGAAGAACTCTACAACTT 252
Qy 81 PheGlyValLysLysIleAArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 253 TACAAGGTTAAGACCATCATCGTGTGGCATTGGCGCAGATTGATCCAAATGTTTAC 312
Qy 101 LeuArgAspValValIleGlyMetCysAlaCysThrAspSerLysValAsnArgIleArg 120
Db 313 GTCCGCGATGTCGTCATTTGTCACAGCATCTGGCAGATTCAAATGTTTAAACAGATGGC 372
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 373 CTTCTTGCCACGATTTCCAGCTACAGCCAACTTCGAAGTCGTTAGTCCTTAGTTGAA 432
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 433 TCCGCCAAGGCTCTCAACATCCCAACACAGGTTGGAAAGGCTACTCAACAGATATCTTC 492
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 493 TACAGCAAGGAACAGGATTAATAGAGCCCTTGCTCAGTACCACCTTCATTGCTGTCGA 552
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 553 ATGGAATCTGCTGTCCTCCCAATCTGATTAATGCGCAAGAGCTGGCTGCATC 612
Qy 201 CysThrValSerAspHisIleAArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 613 TGCACAGTTTCGGATCATCATCACAGTCCGCTACACAGGCTTACTCAACAGATATCTTC 672

RESULT 6
CV205285 794 bp mRNA linear EST 16-SEP-2004
LOCUS EST864995 non-normalized T1 cDNA library Trichomonas vaginalis cDNA
DEFINITION clone TVTAC60 5' end, mRNA sequence.
ACCESSION CV205285
VERSION CV205285.1 GI:52152265
KEYWORDS EST.
SOURCE Trichomonas vaginalis
ORGANISM Trichomonas vaginalis
Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
Trichomonadidae; Trichomonadinae; Trichomonas.
1 (bases 1 to 794)
Carlton,J.M., Dyll,S., Johnson,P.J. and Fraser,C.M.
The complete genome sequence of the sexually transmitted parasite
Trichomonas vaginalis
Unpublished (2004)
Other ESTs: EST864994
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
Seq primer: lambda Triplex2.
Location/Qualifiers
1..794
/organism="Trichomonas vaginalis"
/mol_type="mRNA"
/strain="T1"
/db_xref="taxon:5722"
/clone_lib="non-normalized T1 cDNA library"
/notes="Vector: Lambda Triplex2; Site 1: SfiI; Site 2:
SfiB; T. vaginalis strain T1 library constructed from
cDNA, made in lambda Triplex2. Inserts cloned
unidirectionally in the SfiI and SfiB sites. Mass excision
of library produced inserts in pTriplex2 plasmid. Inserts
```

sequenced from both 5' and 3' ends using Triplex2
sequencing primer and polydtr 24 bp primer respectively."

```
ORIGIN
Alignment Scores: 3.3e-75 Length: 794
Pred. No.: 681.00 Matches: 134
Score: 75.21% Conservative: 42
Percent Similarity: 75.21% Mismatches: 58
Best Local Similarity: 57.26% Indels: 0
Query Match: 55.59% Gaps: 0
DB: 7
US-10-035-300A-2 (1-239) x CV205285 (1-794)
Qy 1 MetAlaThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetPro 20
Db 13 ATGGCAACACCCCACTCTGCAAGGTTGCGGATTTTCGCTGAGACAGTTTCTCATGTGC 72
Qy 21 GlyAspProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluVal 40
Db 73 GCGCATCCACTCCGTCGCCAAGCTCATTTGCTGATACTATCTTGAGAACGCTAAGCAAGTC 132
Qy 41 AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerVal 60
Db 133 AACTCTGTTCTGTGAATGCTTTGGTTTACAGGCACATACAAGGCGCAACCCACTCTCTGTT 192
Qy 61 MetGlyHisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAsp 80
Db 193 ATGGGTATGGCATGGGTATCCCATTTCCATCTACGCTGAAGAACTCTACAACTGTT 252
Qy 81 PheGlyValLysLysIleAArgValGlySerCysGlyAlaValLeuProHisValLys 100
Db 253 TACAAGGTTAAGACCATCATCGTGTGGCATTGGCGCAGATTGATCCAAATGTTTCCAC 312
Qy 101 LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg 120
Db 313 GTCCGCGATGTCGTCATTTGTCACAGCATCTGGCAGATTCAAATGTTTAAACAGATAGCGC 372
Qy 121 PheLysAspHisAspPheAlaIleAlaAspPheAspMetValArgAsnAlaValAsp 140
Db 373 CTTCTTGCCACGATTTCCAGCTACAGCCAACTTCGAAGTCGTTAGTCCTTAGTTGAA 432
Qy 141 AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe 160
Db 433 TCCGCCAAGGCTCTCAACATCCCAACACAGGTTGGAAAGGCTTACTCAACAGATATCTTC 492
Qy 161 TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeuGlyValGlu 180
Db 493 TACAGCAAGGAACAGGATTAATGAGGCGCTTGTCTAGTACCACCTTCATTGCTGTCGAA 552
Qy 181 MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle 200
Db 553 ATGGAATCTGCTGTCCTCCCAATTTGCTGATTAATGCGCAAGAGCTGGCTGCATC 612
Qy 201 CysThrValSerAspHisIleAArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr 220
Db 613 TGCACAGTTTCGGATCATCATCACAGTCCGCTACACAGGCTTACTCAACAGATATCTTC 672
Qy 221 ThrPheAsnAspMetIleLysIleAlaLeuGluSerValLeu 234
Db 673 TCATTCCAGAACATGATCAAGATCGCTCTTGAAGCTACATTG 714
```

```
RESULT 7
CV205285 721 bp mRNA linear EST 20-JUL-2004
LOCUS TVEST085E09 TV30236_PT cDNA Library Trichomonas vaginalis cDNA 5',
DEFINITION mRNA sequence.
ACCESSION CV205285
VERSION CV205285.1 GI:50408335
KEYWORDS EST.
SOURCE Trichomonas vaginalis
ORGANISM Trichomonas vaginalis
Eukaryota; Parabasalidea; Trichomonada; Trichomonadida;
```


	Qy	AsnAsnValArgGlyMetLeuLeuGlyPheThrGlyThrTyrIysGlyAryGlysIleSerVal	60
	Db	 AAACCAAGTTCCGGAAATCTTGCGTTACTTCGAACTTACAAGGGCCATCGTGTTCCTGTA	467
	Qy	MetGlyHisGlyValGlyIleProSerCysSerileTyThrIysGluleuIleThrAsp	80
	Db	 ATGGGAACAGGAATTGGGAATCTTCATACTCATGTATGAATGATTAACAGA	407
	Qy	PheGlyVallYstYsIlelleArgValGlySerCysGlyAlaValLeuProHisValLys	100
	Db	 TATGGCGTTAAAAGATTAAATTCGTGTGGAGCTGCTGTTCTTGTAATGAAGATGTTTCAT	347
	Qy	LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerIysValAsnArgIleArg	120
	Db	 ATTGCGCACCTTGCATTTGGTCAGCGACGACCAACACTTCAGCCATTTGCCGTATCAT	287
	Qy	PheLysAspHisAspPheAlaalaIleAlaAspPheAspMetValArgAsnAlaValaAsp	140
	Db	 TTTTCTGATTTTGATTTCCCACAATTTGCTGATTTTGATTTGCTTGATAAGCTTATCAT	227
	Qy	AlaAlalysAlaLeuGlyIleAspAlaargValGlyAsnLeuPheSerAlaAspLeuPhe	160
	Db	 ATTGCMAAAGACCCTTGGAATTTACAACCTCACCGCTGGGAATATTTCTAGTTCCTGATTGTTC	167
	Qy	Tyr---SerProAspGlyGluMetPheAspValMetGluLylsTyzGlylIleuGlyVal	179
	Db	 TATGGTGGACCAAGTCAGTTAAA-----GTTGANAACCTCGGTGTGAAGCGCTGTC	116
	Qy	GluMetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPhGelyAlalySAlaleuthr	199
	Db	 GAAATGGAAGCAGCTGCTCTTATTATCTTGGGGCAAATAATAAAGTCCAAACITTTGGGA	56
	Qy	IleCysThrValSerAspHisIleArgThrHisGlu	211
	Db	 ATTATGCAATTTCTGCACCATATTTTTCACAGAGAA	20

```

RESULT 9
  AZ533877/c
  LOCUS
  DEFINITION
  ACCESSION
  VERSION
  KEYWORDS
  SOURCE
  ORGANISM
  REFERENCE
  AUTHORS
  TITLE
  JOURNAL
  COMMENT
  AZ533877
  ENTBJ77R Entamoeba histolytica Sheared DNA Entamoeba histolytica
  genomic, genomic survey sequence.
  AZ533877.1 GI:11089945
  GSS.
  Entamoeba histolytica
  Entamoeba histolytica
  Eukaryota; Entamoebidae; Entamoeba.
  1 (bases 1 to 878)
  Loftus,B., Van Aken,S. and Fraser,C.
  Determination of clone end sequences from Entamoeba histolytica
  HM1:IMSS sheared DNA library
  Unpublished (2000)
  Contact: Brendan J Loftus
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850, USA
  Tel: 301 838 0208
  Fax: 301 838 3543
  Email: bjloftus@tigr.org
  Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
  DNA library
  Seq primer: M13-Reverse
  Class: shotgun
  High quality sequence start: 14
  High quality sequence stop: 847.
  Location/Qualifiers
    1..878
      /organism="Entamoeba histolytica"
      /mol_type="genomic DNA"
      /strain="HM1:IMSS"
      /db_xref="taxon:5759"
      /clone_lib="Entamoeba histolytica Sheared DNA"
  FEATURES
    source

```

/note=Vector: pHO51; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

Alignment Scores:		
Prod. No.:	3_23e-44	Length: 878
Score:	435.50	Matches: 88
Percent Similarity:	63.68%	Conservative: 33
Best Local Similarity:	46.32%	Mismatches: 68
Query Match:	35.63%	Indels: 1
DB:	8	Gaps: 1

US-10-035-300A-2 (1-239) x A2533877 (1-878)

QY	41	AsnAsnValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyArgLysIleSerVal	60
DB	878	AATGGAGTAAGAGGGCATTAGATATATGATATACAAAGAAAAGGTTCTATC	819
QY	61	MetGlyHisGlyValGlyLeuProSerCysSerIleTyThrLysGluLeuIleThrAsp	80
DB	818	AAAGCTCATGGTATGGGTATACCATCATTTATGGAAATTTATGCATATGAATTTATTAATTTTC	759
QY	81	PheGlyValLysLysIleIleArgValGlySerCysGlyAlaValLeuProHisValLys	100
DB	758	TATGGAGTAAAGAAATTTATGAAATTTGTTCTGCTGGAGCTATTTAGCTCTAAATTAGAT	699
QY	101	LeuArgAspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArg	120
DB	698	ATTGGAGATATTGTCATAGCAACTGGTGTGTCATATAATAGTAATTTATATGCAACAATAT	639
QY	121	PheLysAspHisAspPheAlaIleAlaAspPheAspMetValargAsnAlaValAsp	140
DB	638	AAACTCCAGGACATCTTCGCCAATAGCTGATATTAGAGTTAGTTCTATAAGCTGTTGAA	579
QY	141	AlaAlaLysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPhe	160
DB	578	GCTGTTGAAAGATTGGAGCTAAGTATGCNAGTAGGAAATGTTTTCTAGTGATGTTTC	519
QY	161	TyrSerProAspGlyGluMetPheAspValMetGluLysTyrGlyLeuGlyValGlu	180
DB	518	TATGAT---GATTCAGTTCATTTAAAAGAATGGATAAGATGGAGTACTTGCTGTAGAA	462
QY	181	MetGluAlaAlaGlyIleTyrGlyValAlaAlaGluPheGlyAlaLysAlaLeuThrIle	200
DB	461	ATGGAAGACGCTGGTTTATGCTATTTCGGCGAAGACAGGCAAAAGCTTTGACTATT	402
QY	201	CysThrValSerAspHisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThr	220
DB	401	TGTTCTATTCTGATGTCCATTTAAGGAACTGCTCTCTCAGCTGAAGAAACAAACC	342
QY	221	ThrPheAsnAspMetIleLysIleAlaLeu	230
DB	341	AAATTTACAAATGATGGAAGTTGCACTG	312

RESULT 10	BH138974	921 bp	DNA	linear	GSS 07-AUG-2001
LOCUS	BH138974				
DEFINITION	ENTOV94FF Entamoeba histolytica Sheared DNA				Entamoeba histolytica
	genomic, genomic survey sequence.				
ACCESSION	BH138974				
VERSION	BH138974.1				GI:15098035
KEYWORDS	GSS.				

SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 921)
 Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 TITLE HMI:IMSS sheared DNA library (2001)
 JOURNAL Unpublished (2001)
 COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 2
 High quality sequence stop: 563.
 FEATURES
 source
 1..921
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /notes="Vector: pHSI; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Barell, Oxford University Press, 1999)."

ORIGIN

Alignment Scores:
 Pred. No.: 3,99e-44 Length: 921
 Score: 436.00 Matches: 108
 Percent Similarity: 65.18% Conservative: 38
 Best Local Similarity: 48.21% Mismatches: 74
 Query Match: 35.59% Indels: 7
 DB: 8 Gaps: 3

US-10-035-300A-2 (1-239) x BH138974 (1-921)

QY 3 ThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAsp 22
 Db 130 ACTCCACATAATGCTGCTAAATACGAGAGATGCTGAAACTGCTCTTATGGCTGGTAT 189
 QY 23 ProLeuArgAlaLysTyrrIleAlaGluThrPheLeuGluAspAlaArgGluValAsn 42
 Db 190 CCATTAAAGATAAACTCTTTCGACAGACACTATTTCGACTGATGTTCAATAATAAGT 249
 QY 43 ValArgGlyMetLeuGlyPheThrGlyThrTyrrLysGlyArgLysIleSerValMetGly 62
 Db 250 GTTAGAGGACGATGAGATATATCTGGATATACAGGAGTGAAGTATACAGTTCAAGCT 309
 QY 63 HisGlyValGlyIleProSerCysSerIleTyrrThryLysGluLeuIleThrAspPheGly 82
 Db 310 CATGGTATGGTATGCCATCTATTGGAAATTTATGCATATCAATATTATTAATTTCTATGA 369
 QY 83 ValLysAlaLysIleLeuArgValGlySerCysGlyAlaValLeuProHisValLysLeuArg 102
 Db 370 GTAAGAGAAATTATTAGAAATGGTTCTGCTGGAGCTTTTCATGAAAGTTTAAACTTGA 429

QY 103 AspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleAArgPheLys 122
 Db 430 GATATGTTTATGGAATGGAGCAATGTTATGATTCCTCAATTTGAAAGACAATATGATATT 489
 QY 123 AspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAla 142
 Db 490 CCAGGTAAATACTCATGCTATTGCTGATTTCCAACTTTGCAGAGAAGCAGTTGATGCAGCT 549
 QY 143 LysAlaLeuGlyIleAspAlaArgValGlyAsnLeuPheSerAlaAspLeuPheTyrSer 162
 Db 550 GAAAACTCGATATAGATAT-AAAAGTAAGAAATATTATTCTCTAATATTATTCTTATGAT 608
 QY 163 ProAspGlyGluMetPheAspValMetGluLysTyrrGlyIleLeuGlyValGluMetGlu 182
 Db 609 --GATGGAGATCACTCCGGAGCATCGAAAAAGATGGGAGTACTTCTGTAGAAAT-GAA 664
 QY 183 AlaAlaGlyIleTyrrGlyValAlaAlaGluPheGlyAlaLys-AlaLeuThrIleCysTh 202
 Db 665 GCAGCTGCTCTTTATATGATTTGCAGCAGACAC--GTAACACAGCACTTTGTATGGTAAC 721
 QY 202 rValSerAsp---HisIleArgThrHisGluGlnThrThrAlaAlaGluArgGlnThrTh 221
 Db 722 TATTTCTGATCTTTGTTATGGAAGTGGGAAAAATG-ACAGCCGAAAGAAAGAACTAAT 780
 QY 221 rPheAsnAsp 224
 Db 781 TTTCTCCAAT 790

RESULT 11
 AZ680406/c

LOCUS ENTGM406F Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000
 DEFINITION genomic, genomic survey sequence.

ACCESSION AZ680406
 VERSION AZ680406.1 GI:11817552
 KEYWORDS GSS.

SOURCE

Entamoeba histolytica

Entamoeba histolytica

Eukaryota; Entamoebidae; Entamoeba.

1 (bases 1 to 863)

Loftus, B., Van Aken, S. and Fraser, C.

Determination of clone end sequences from Entamoeba histolytica

HMI:IMSS sheared DNA library

Unpublished (2000)

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 26

High quality sequence stop: 769.

FEATURES

source

1..863
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /notes="Vector: pHSI; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

ORIGIN

Alignment Scores:
 Pred. No.: 2,428-43 Length: 863
 Score: 429.50 Matches: 83
 Percent Similarity: 65.14% Conservative: 31
 Best Local Similarity: 47.43% Mismatches: 60
 Query Match: 35.06% Indels: 1
 DB: 8 Gaps: 1

US-10-035-300A-2 (1-239) x AZ680406 (1-863)

Qy 3 ThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAsp 22
 Db 545 ACTCCACACACACGCTAATATAGCTGATTTGCTGATACAGTGTGTAGCTGGTAT 486

Qy 23 ProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsnAsn 42
 Db 485 CCACCTAAGATGTAATTTGTTGCACACACTTATTTGACTGATGTTGTTCAAGTCAATGGA 426

Qy 43 ValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMetGly 62
 Db 425 GTAAGAGGGCATTAGGATATAGTATGATATACAAAGGAAAGGTTTCTATTCAAGCT 366

Qy 63 HisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPheGly 82
 Db 365 CATGGTATGGGTATACCATCTATTTGGAATTTATGATGATGATGATTAATTTCTATGGA 306

Qy 83 VallysLysIleIleArgValGlySerCysGlyAlaValLeuProHisVallysLeuArg 102
 Db 305 GTAAGAGAAATTTAGAAATGGTTCTGCTGGAGTATAGCTCTAAATTTAGATTTGGA 246

Qy 103 AspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLys 122
 Db 245 GATATTGTATAGCACTGGTCTGCTGATATAGTATATTTTATGCAACAATATAAATCT 186

Qy 123 AspHisAspPheAlaAlaIleAlaAspPheMetValArgAsnAlaValAspAlaAla 142
 Db 185 CCAGGAACATTTGCGCCAAATAGCTGATTTAGTATGTTCTATAAAGCTGTTGAAGCT 126

Qy 143 LysAlaLeuGlyIleAspAlaArgValGlyAlaLeuPheSerAlaAspLeuPheTyrSer 162
 Db 125 GAAAGATGGAGCTAAGTATGCTAGTGGATGTTTCTTCTAGTATGTTCTATGAT 66

Qy 163 ProAspGlyGluMetPheAspValMetGluLysTyrGlyIleLeu 177
 Db 65 ---GATTCAAGTTTCATTAAGAATGGGATAGGATGAGTACTT 24

RESULT 12
 BH166036/c 861 bp DNA linear GSS 24-SEP-2001
 LOCUS ENT5Y60TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.

ACCESSION BH166036
 VERSION BH166036.1 GI:15739474
 KEYWORDS GSS.
 ORGANISM Entamoeba histolytica
 SOURCE Entamoeba histolytica
 ORGANISM Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE 1 (bases 1 to 861)
 Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 TITLE HMI:IMSS sheared DNA library (2001)
 JOURNAL Unpublished (2001)

COMMENT
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208

Fax: 301 838 3543
 Email: bjloftus@igr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library

FEATURES

source
 1. 861
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHS1; Site 1: Bet I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 using a method described by Clark and Diamond (Clark,
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Borell, Oxford University Press, 1999)."

ORIGIN

Alignment Scores:
 Pred. No.: 1,428-36 Length: 861
 Score: 376.00 Matches: 71
 Percent Similarity: 68.09% Conservative: 25
 Best Local Similarity: 50.35% Mismatches: 45
 Query Match: 30.69% Indels: 0
 DB: 8 Gaps: 0

US-10-035-300A-2 (1-239) x BH166036 (1-861)

Qy 3 ThrProHisIleAsnAlaGluMetGlyAspPheAlaAspValValLeuMetProGlyAsp 22
 Db 464 ACTCCACACACACGCTAATATAGCTGATTTGCTGATACAGTGTGTAGCTGGTAT 405

Qy 23 ProLeuArgAlaLysTyrIleAlaGluThrPheLeuGluAspAlaArgGluValAsnAsn 42
 Db 404 CCACCTAAGATGTAATTTGTTGCACACACTTATTTGACTGATGTTGTTCAAGTCAATGGA 345

Qy 43 ValArgGlyMetLeuGlyPheThrGlyThrTyrLysGlyArgLysIleSerValMetGly 62
 Db 344 GTAAGAGGGCATTAGGATATAGTATGATATACAAAGGAAAGGTTTCTATTCAAGCT 285

Qy 63 HisGlyValGlyIleProSerCysSerIleTyrThrLysGluLeuIleThrAspPheGly 82
 Db 284 CATGGTATGGGTATACCATCTATTTGGAATTTATGATGATGATTAATTTCTATGGA 225

Qy 83 VallysLysIleIleArgValGlySerCysGlyAlaValLeuProHisVallysLeuArg 102
 Db 224 GTAAGAGAAATTTAGAAATGGTTCTGCTGGAGTATAGCTCTAAATTTAGATTTGGA 165

Qy 103 AspValValIleGlyMetGlyAlaCysThrAspSerLysValAsnArgIleArgPheLys 122
 Db 164 GATATTGTATAGCACTGGTCTGCTGATATAGTATATTTTATGCAACAATATAAATCT 105

Qy 123 AspHisAspPheAlaAlaIleAlaAspPheMetValArgAsnAlaValAspAlaAla 142
 Db 104 CCAGGAACATTTGCGCCAAATAGCTGATTTAGTATGTTCTATAAAGCTGTTGAAGCT 45

Qy 143 Lys 143
 Db 44 AAA 42

RESULT 13

```

AZ675191
LOCUS      AZ675191      864 bp      DNA      linear      GSS 14-DEC-2000
DEFINITION Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION  AZ675191
VERSION    AZ675191.1  GI:11812337
KEYWORDS   GSS.
SOURCE     Entamoeba histolytica
            Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 864)
AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
TITLE     Determination of clone end sequences from Entamoeba histolytica
            HML:IMSS sheared DNA library
JOURNAL   Unpublished (2000)
COMMENT   Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjltoftus@tigr.org
            Clones are derived from the Entamoeba histolytica HML:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 21
            High quality sequence stop: 808.

FEATURES             source
     source
     1..864
         /organism="Entamoeba histolytica"
         /mol_type="genomic DNA"
         /strain="HML:IMSS"
         /db_xref="taxon:5759"
         /clone_lib="Entamoeba histolytica Sheared DNA"
         /note="Vector: pHOS1; Site 1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999)."
```

```

         /organism="Entamoeba histolytica"
         /mol_type="genomic DNA"
         /db_xref="taxon:5759"
         /clone_lib="Entamoeba histolytica Sheared DNA"
         /note="Vector: pHOS1; Site 1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barrell, Oxford University Press, 1999)."
```

589 CATGGTATGGTATGCCATCTATTGGAAATTTATGCATATGAATTTAATTTCTATGGA 648

83 Vallyalysllelearg-ValglySerCysGlyAlaValLeuProHisVallylsleuAr 102

649 GTAAAGAGAAATATTAGAAATTCGTTCTGCTGAGGCTTTGATGAAAGTTTAAACATGG 708

102 GAspValValleleGlyMetGlyAlaCysThrAspSerLysValAsnArglleargPheLy 122

709 AGATATTGTTATTGGAAATGGAGCATGTTATGATTTCCAAATTTGAAAGACAATATGATAT 768

122 sAspHisAspPheAlaAlaIleAlaAspPheAspMetValArgAsnAlaValAspAlaAl 142

769 TCCAGGTAAATACTCATGTATTGCTGATTTCCAACTTTGCAGAGAAACAGTTGATGCAGC 828

142 aLysAlaLeuGlylleAspAlaArgValGlyAsnLeu 154

829 TGAATAA-ACCTGGTATATAGATATAAGGTTAGGAAATTTA 864

RESULT 14

AZ669801/c

LOCUS AZ669801 865 bp DNA linear GSS 14-DEC-2000

DEFINITION ENTLL39TF Entamoeba histolytica Sheared DNA Entamoeba histolytica

genomic, genomic survey sequence.

ACCESSION AZ669801

VERSION AZ669801.1 GI:11806947

KEYWORDS GSS.

SOURCE Entamoeba histolytica

Entamoeba histolytica

Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 865)

AUTHORS Loftus,B., Van Aken,S. and Fraser,C.

TITLE Determination of clone end sequences from Entamoeba histolytica

HML:IMSS sheared DNA library

JOURNAL Unpublished (2000)

COMMENT Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjltoftus@tigr.org

Clones are derived from the Entamoeba histolytica HML:IMSS sheared

DNA library

Seq primer: M13-Forward

Class: shotgun

High quality sequence start: 18

High quality sequence stop: 707.

FEATURES source

source

1..865

/organism="Entamoeba histolytica"

/mol_type="genomic DNA"

/strain="HML:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: pHOS1; Site 1: Bst I; Constructed at The

Institute for Genomic Research (TIGR), Rockville, MD.

Genomic DNA isolated from broth cultures of E. histolytica

using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a

method for isolate identification. Exp. Parasitol.

77:450.). The DNA was mechanically sheared to give a

tight size distribution (~2 kb). The v + i method used for

the library construction is described in detail in Smith,

H.O. and Venter, J.C. (Making small insert libraries for

whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.

Barrell, Oxford University Press, 1999)."

ORIGIN

Alignment Scores: 1.84e-32 Length: 865

Pred. No.: 343.50 Matches: 64

Score:

Percent Similarity: 68.95% Conservative: 20
Best Local Similarity: 52.46% Mismatches: 33
Query Match: 28.04% Indels: 5
DB: 8 Gaps: 1

US-10-035-300A-2 (1-239) x AZ681446 (1-865)

Qy 3 ThrProHisIleAenAlaGluMetGlyAaspPheAlaAaspValValLeuMetProGlyAasp 22
Db 365 ACTCCACACAAACGCTAAATAGCTGATTGCTGATACAGTGTGTAGTCTGGTGTAT 306
Qy 23 ProLeuArgAlaLysTyrlleAlaGluThrPheLeuGluAaspAlaAargGluValAasn 42
Db 305 CCACCTAAGATGTAATTTGTTGCAGACACTTATTGCTGATGTTGTTCAAGTCAATGGA 246
Qy 43 ValArgGlyMetLeuGlyPheThrGlyThrTyrlsGlyAargLysIleSerValMetGly 62
Db 245 GTAGAGGGGATAGGATATCTGATATTAACAAGGAAAAAGTTTCTATTCAAGCT 186
Qy 63 HisGlyValGlyIleProSerCysSerIleTyrlsGlyGluLeuIleThrAaspPheGly 82
Db 185 CATGGTATGGGTATACCATCTATTGGAATTTATGATGATGATGATTTAAATTTCTATGGA 126
Qy 83 ValLysIleIleAargValGlySerCysGlyAlaValLeuProHisValLysLeuArg 102
Db 125 GTAAAGAGAAATTTAGAAATTTGCTGCTGAGCTATTAGCTCTAAATTAGATATTGGA 66
Qy 103 AspValValIleGlyMetGlyAlaCysThrAaspSerLysValAasnArgIleAargPheLys 122
Db 65 GATATTGTCATAGCACTGGTGTGCTGATATAGT-----AATTTATG 21
Qy 123 AspHis 124
Db 20 CAACAT 15

RESULT 15
AZ681446/c
LOCUS
DEFINITION ENTMK91TR Entamoeba histolytica Sheared DNA linear GSS 14-DEC-2000
genomic, genomic survey sequence.
ACCESSION AZ681446
VERSION AZ681446.1 GI:11818592
KEYWORDS GSS.
SOURCE Entamoeba histolytica
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 865)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
Unpublished (2000)
CONTACT: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 14
High quality sequence stop: 804.
Location/Qualifiers
1. .882
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.

FEATURES

source
1. .882
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.

ORIGIN

Alignment Scores:
Pred. No.: 8.82e-29 Length: 882
Score: 314.50 Matches: 68
Percent Similarity: 60.38% Conservative: 28
Best Local Similarity: 42.77% Mismatches: 62
Query Match: 25.67% Indels: 1
DB: 8 Gaps: 1
US-10-035-300A-2 (1-239) x AZ681446 (1-882)
Qy 72 IleTyrlsThrLysGluLeuIleThrAaspPheGlyValLysLysIleIleAargValGlySer 91
Db 876 ATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 817
Qy 92 CysGlyAlaValLeuProHisValLysLeuAargValValIleGlyMetGlyValaCys 111
Db 816 GCTGAAGCTATAGCTCTAAATAGATATGAGATATGTCATAGCACTGGTGTGT 757
Qy 112 ThrAaspSerLysValAasnArgIleAargPheLysAaspPheAlaIleAlaAasp 131
Db 756 CATATAGTATATTTATGCAACAATATAAAGCTCCAGGAACATTTGCCCAATAGCTGAT 697
Qy 132 PheAaspMetValArgAasnAlaValAaspAlaLysAlaLeuGlyIleAaspAlaArgVal 151
Db 696 TATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 637
Qy 152 GlyAasnLeuPheSerAlaAaspLeuPheTyrlsSerProAaspGlyGluMetPheAaspValMet 171
Db 636 GGGAAATGTTTCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 580
Qy 172 GlulysTyrlsGlyIleLeuGlyValGluMetGluAlaAlaGlyIleTyrlsGlyValaAla 191
Db 579 GATAAGATGGAGTACTTGTCTGATAGAAATGAAGCAGCTGCTGATATGATGATGATGATGAT 520
Qy 192 GluPheGlyAlaLysAlaLeuThrIleCysThrValSerAaspHisIleAargThrHisGlu 211
Db 519 AGACGAGGCAACAAAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 460
Qy 212 GlnThrThrAlaAlaGluAargGlnThrThrPheAasnAaspMetIleLysIleAlaLeu 230
Db 459 GCTCTCTCAGCTGAGAGAAAGCAACCAATTTACACAAATGATGAGAGTTGCATG 403

Search completed: July 23, 2005, 04:25:30
Job time : 3882 secs

This Page Blank (Inten)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 2005, 21:22:05 ; Search time 174 Seconds
(without alignments)
703.373 Million cell updates/sec

Title: US-10-035-300A-2
Perfect score: 1225
Sequence: 1 MATPHINAEMGDFADVLMF.....TTFNDMIKIALESVLLGDKE 239

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03.4

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1221	99.7	239	Q7UA16	Q7ua16 shigella fl
2	1221	99.7	258	Q83P00	Q83p00 shigella fl
3	1221	99.7	258	Q8FA51	Q8fa51 escherichia
4	1217	99.3	238	1 DEOD_ECOLI	P09743 escherichia
5	1187	96.9	239	2 Q8Z0F2	Q8z0u2 salmonella
6	1182	96.5	239	2 Q8ZJV7	Q8zjv7 salmonella
7	1115	91.0	239	2 Q66EV7	Q66ev7 yersinia ps
8	1106	90.3	239	2 Q8ZIQ2	Q8zic2 yersinia pe
9	1106	90.3	241	2 Q8CZM9	Q8czn9 yersinia pe
10	1095	89.4	239	2 Q6D989	Q6d989 erwinia car
11	1085	88.7	238	2 Q8KRT5	Q8krt5 xenorhabdus
12	1085	88.6	239	1 DEOD_KLEPN	Q59482 klebsiella
13	1067	87.1	238	2 Q7N930	Q7n930 photorhabdu
14	1039	84.8	238	2 Q7N930	Q7n930 photorhabdu
15	1021	83.3	238	2 Q65RA4	Q65ra4 manheimia
16	1018	83.1	239	2 Q87M25	Q87m25 vibrio para
17	1017	83.0	238	1 DEOD_HAEIN	P44417 haemophilus
18	1012	82.6	239	2 Q8DE59	Q8db59 vibrio vuln
19	1012	82.6	254	2 Q7MI41	Q7mi41 vibrio vuln
20	996	81.3	241	2 Q6LUH1	Q6luh1 photobacter
21	994	81.1	240	1 DEOD_ACTPL	P94164 actinobacil
22	993	81.1	241	2 Q9KPM0	Q9kpm0 vibrio chol
23	993	76.2	238	2 Q7VMS8	Q7vms8 haemophilus
24	861	70.3	236	2 Q8EKK0	Q8ekko shewanella
25	858	70.0	236	2 Q6LLA7	Q6lla7 photobacter
26	858	70.0	245	2 Q9KNE2	Q9knb2 vibrio chol
27	853	69.6	236	2 Q87G42	Q87g42 vibrio para
28	834	68.1	236	2 Q8D322	Q8d322 vibrio vuln
29	834	68.1	274	2 Q7MFC6	Q7mfc6 vibrio vuln
30	778	63.5	234	1 DEOD_BUCAI	P57606 buchnera ap
31	747	61.0	236	2 Q7NRT2	Q7nrt2 chromobacte

32	729	59.5	236	2 Q8EKK0	Q8ekko shewanella
33	712	58.1	238	1 DEOD_BUCBP	Q89a58 buchnera ap
34	700	57.1	235	2 Q63DR9	Q63dr9 bacillus ce
35	699	57.1	235	2 Q73B32	Q73b32 bacillus ce
36	699	57.1	235	2 Q81T09	Q81t09 bacillus an
37	699	57.1	235	2 Q6HL92	Q6hl92 bacillus th
38	698	57.0	235	2 Q81FV5	Q81fv5 bacillus ce
39	693.5	56.6	233	1 DEOD_BACSU	Q34925 bacillus su
40	690	56.3	236	1 DEOD_BUCAP	Q8k937 buchnera ap
41	675.5	55.1	235	2 Q8ENY0	Q8eny0 oceanobacil
42	675	55.1	232	2 Q8R973	Q8r973 thermoanaer
43	675	55.1	233	1 DEOD_HELPY	P56463 helicobacte
44	675	55.1	305	2 Q9RSF8	Q9rsf8 deinococcus
45	671.5	54.8	233	2 Q92AF2	Q92af2 listeria in

ALIGNMENTS

RESULT 1

Q7UA16	ID	Q7UA16	PRELIMINARY;	PRT;	239 AA.
AC	Q7UA16;	01-OCT-2003 (T-EMBLrel. 25, Created)			
DT	01-OCT-2003 (T-EMBLrel. 25, Last sequence update)				
DE	01-MAR-2004 (T-EMBLrel. 26, Last annotation update)				
DE	Purine-nucleoside phosphorylase.				
GN	Name=deod; OrderedLocusNames=S4687;				
OS	Shigella flexneri.				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Shigella.				
OX	NCBI_TaxID=623;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=2457T;				
RX	MEDLINE=22590274; PubMed=12704152;				
RX	DOI=10.1128/JAI.71.5.2775-2786.2003;				
RA	Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,				
RA	Fournier G., Mayhew G.P., Plunkett G. III, Rose D.J., Darling A.,				
RA	Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,				
RA	Schwartz D.C., Blattner F.R.;				
RT	"Complete genome sequence and comparative genomics of Shigella				
RT	flexneri serotype 2a strain 2457T."				
RL	Infect. Immun. 71:2775-2786(2003).				
DR	EMBL; AE016993; AAP19605.1; -.				
DR	HSP; P09743; IK9S.				
DR	GO; GO:0003824; F:catalytic activity; IEA.				
DR	GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.				
DR	GO; GO:0009116; F:nucleoside metabolism; IEA.				
DR	InterPro; IPR004402; Deod.				
DR	Pfam; PF01048; PNP_UDP_1; 1.				
DR	TIGRFAMS; TIGR00107; deod; 1.				
DR	PROSITE; PS01232; PNP_UDP_1; 1.				
SQ	SEQUENCE 239 AA; 25936 MW; FAS0743A44P87781 CRC64;				

Query Match 99.7%; Score 1221; DB 2; Length 239;
Best Local Similarity 99.2%; Pred. No. 2.3e-92;
Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MATPHINAEMGDFADVLMF	QYAKYIAETFLDAREVNNVGM	GFTGTYGKRISV	60
Db	1	MATPHINAEMGDFADVLMF	QYAKYIAETFLDAREVNNVGM	GFTGTYGKRISV	60
Qy	61	MGHGVGPSCSYITKELITD	FGVKKIIRVSCGAVLPHVKLRD	VVIGWACTDSKVNIR	120
Db	61	MGHGVGPSCSYITKELITD	FGVKKIIRVSCGAVLPHVKLRD	VVIGWACTDSKVNIR	120
Qy	121	FKQHDFAAIADFMVRNAVDA	AKALGIDARVGNLFSADLFYSP	PDGEMFDMVEKYGILGVE	180
Db	121	FKQHDFAAIADFMVRNAVDA	AKALGIDARVGNLFSADLFYSP	PDGEMFDMVEKYGILGVE	180
Qy	181	MEAAAGIVGVAEFGAKALT	TTCTVSDHRTTHEQTAAERQT	TTFNDMIKIALESVLLGDKE	239

```
Db 181 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTTFNDMIKIALESVLLGDKE 239
|||||
RESULT 2
Q83P00 PRELIMINARY; PRT; 258 AA.
AC Q83P00;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Purine-nucleoside phosphorylase.
GN Name=deob; OrderedLocusNames=SP4416;
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RC MEDLINE=22272406; PubMed=12384590; DOI=10.1093/nar/gkf566;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441 (2002).
DR EMBL; AE015447; AAN45831.1; -.
DR HSSP; P09743; 1K9S.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deob.
DR InterPro; IPR000845; PNP_UDP.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deob; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 258 AA; 27956 MW; 3E0BCAA959374365 CRC64;

Query Match 99.7%; Score 1221; DB 2; Length 258;
Best Local Similarity 99.2%; Pred. No. 2.5e-92;
Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db 20 MATPHINAEMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 79
|||||
QY 61 MGHGVGPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 80 MGHGMGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVLVGMGACTDSKNRIR 139
|||||
QY 121 FKDHFAAIAADFDMVNNVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
Db 140 FKDHFAAIAADFDMVNNVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 199
|||||
QY 181 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTTFNDMIKIALESVLLGDKE 239
Db 200 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTTFNDMIKIALESVLLGDKE 258
|||||

RESULT 3
Q8FA51 PRELIMINARY; PRT; 258 AA.
AC Q8FA51;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1).
GN Name=deob; OrderedLocusNames=c5468;
OS Escherichia coli O6.

Db 181 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTTFNDMIKIALESVLLGDKE 239
|||||
Db 200 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTTFNDMIKIALESVLLGDKE 258
|||||

RESULT 4
DEOB_ECOLI STANDARD; PRT; 238 AA.
AC P09743;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase) (PNP).
DE Name=deob; Synonyms=pup; OrderedLocusNames=b4384, z5986, ECs5343;
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RC MEDLINE=91334430; PubMed=1714590;
RA Hersfield M.S., Chaffee S., Koro-Johnson L., Mary A., Smith A.A.,
RA Short S.A.;
RT Use of site-directed mutagenesis to enhance the epitope-shielding
RT effect of covalent modification of proteins with polyethylene
RT glycol."
RT glycol."
RN Proc. Natl. Acad. Sci. U.S.A. 88:7185-7189 (1991).
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RC MEDLINE=22388234; PubMed=12471157; DOI=10.1073/pnas.252529799;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
DR EMBL; AE016772; AAN83888.1; -.
DR HSSP; P09743; 1K9S.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deob.
DR InterPro; IPR000845; PNP_UDP.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deob; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
DR Complete proteome; Glycosyltransferase; Transferase.
KW Complete proteome; Glycosyltransferase; Transferase.
SQ SEQUENCE 258 AA; 27942 MW; EC05C8DFAB3742C1 CRC64;

Query Match 99.7%; Score 1221; DB 2; Length 258;
Best Local Similarity 99.2%; Pred. No. 2.5e-92;
Matches 237; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MATPHINAEMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db 20 MATPHINAEMGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 79
|||||
QY 61 MGHGVGPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVLVGMGACTDSKNRIR 120
Db 80 MGHGMGIPSCSIYTKELITDFGVKKIIRVSGCAVLPHVKLRDVLVGMGACTDSKNRIR 139
|||||
QY 121 FKDHFAAIAADFDMVNNVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 180
Db 140 FKDHFAAIAADFDMVNNVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVE 199
|||||
QY 181 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTTFNDMIKIALESVLLGDKE 239
Db 200 MEAAGIYGVAAEFAGAKALITCTVSDHIRTHTQTTAAERQTTTFNDMIKIALESVLLGDKE 258
|||||

RESULT 4
DEOB_ECOLI STANDARD; PRT; 238 AA.
AC P09743;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase) (PNP).
DE Name=deob; Synonyms=pup; OrderedLocusNames=b4384, z5986, ECs5343;
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RC MEDLINE=91334430; PubMed=1714590;
RA Hersfield M.S., Chaffee S., Koro-Johnson L., Mary A., Smith A.A.,
RA Short S.A.;
RT Use of site-directed mutagenesis to enhance the epitope-shielding
RT effect of covalent modification of proteins with polyethylene
RT glycol."
RT glycol."
RN Proc. Natl. Acad. Sci. U.S.A. 88:7185-7189 (1991).
```

[2]
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the *Escherichia coli* genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EHEC;
RX MEDLINE=21074935; PubMed=11205551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postfai G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
RA Postbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouzis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
RL Nature 409:529-533(2001).
[4]
RN SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / Sakai / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
[5]
RN SEQUENCE OF 224-238 FROM N.A.
RP STRAIN=K12;
RX MEDLINE=87259990; PubMed=3299264;
RA Larsen J.E.L., Albrechtsen B., Valentin-Hansen P.;
RT "Analysis of the terminator region after the *deoCABD* operon of
RT *Escherichia coli* K-12 using a new class of single copy number operon-
RT fusion vectors.";
RL Nucleic Acids Res. 15:5125-5140(1987).
[6]
RN SEQUENCE OF 1-20.
RP MEDLINE=93281685; PubMed=8506346;
RA Henzel W.J., Billicci T.M., Stults J.T., Wong S.C., Grimley C.,
RA Watanabe C.;
RT "Identifying proteins from two-dimensional gels by molecular mass
RT searching of peptide fragments in protein sequence databases.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:5011-5015(1993).
[7]
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RP MEDLINE=98035458; PubMed=9351810; DOI=10.1016/S0969-2126(97)00287-6;
RA Mao C., Cook W.J., Zhou M., Koszalka G.W., Krenitsky T.A.,
RA Ealick S.E.;
RT "The crystal structure of *Escherichia coli* purine nucleoside
RT phosphorylase: a comparison with the human enzyme reveals a conserved
RT topology.";
RL Structure 5:1373-1383(1997).
[8]
RN X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RP MEDLINE=98317045; PubMed=9653038; DOI=10.1006/jmbi.1998.1799;
RA Koellner G., Luic M., Shugar D., Saenger W., Bzowska A.;
RT "Crystal structure of the ternary complex of *E. coli* purine nucleoside
RT phosphorylase with formycin B, a structural analogue of the substrate
RT inosine, and phosphate (sulphate) at 2.1-A resolution.";
RL J. Mol. Biol. 280:153-166(1998).
CC -!- FUNCTION: Cleavage of guanosine or inosine to respective bases and
CC sugar-1-phosphate molecules.
CC -!- CATALYTIC ACTIVITY: Purine nucleoside + phosphate = purine +
CC alpha-D-ribose 1-phosphate.
CC -!- SUBUNIT: Homohexamer.
CC -!- SIMILARITY: Belongs to the PNP/UDP phosphorylase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL; M60917; AAA24401.1; -;
DR EMBL; U14003; AAA97280.1; -;
DR EMBL; U00096; AAC77337.1; -;
DR EMBL; AE005669; AAC59565.1; -;
DR EMBL; AP002569; BAB38766.1; -;
DR PIR; X05629; CAA29114.1; -;
DR PIR; A41143; A27854.
DR PIR; A86138; A86138.
DR PIR; G91296; G91296.
DR PDB; 1A69; X-ray; A/B/C=1-238.
DR PDB; 1ECP; X-ray; A/B/C/D/E/F=1-237.
DR PDB; 1K9S; X-ray; A/B/C/D/E/F=1-238.
DR PDB; 1OTX; X-ray; A/B/C=1-238.
DR PDB; 1OTY; X-ray; A/B/C=1-238.
DR PDB; 1OU4; X-ray; A/B/C=1-238.
DR PDB; 1OUM; X-ray; A/B/C=1-238.
DR PDB; 1OV6; X-ray; A/B/C=1-238.
DR PDB; 1OVG; X-ray; A/B/C=1-238.
DR PDB; 1PK7; X-ray; A/B/C=1-237.
DR PDB; 1PK9; X-ray; A/B/C=1-237.
DR PDB; 1PKE; X-ray; A/B/C=1-237.
DR PDB; 1PR0; X-ray; A/B/C=1-238.
DR PDB; 1PR1; X-ray; A/B/C=1-238.
DR PDB; 1PR2; X-ray; A/B/C=1-238.
DR PDB; 1PR4; X-ray; A/B/C=1-238.
DR PDB; 1PR5; X-ray; A/B/C=1-238.
DR PDB; 1PR6; X-ray; A/B/C=1-238.
DR PDB; 1PW7; X-ray; A/B/C=1-238.
DR SWISS-2DPAGE; P09743; COLI.
DR EcoGene; EG1022; deoD.
DR InterPro; IPR004402; deoD.
DR InterPro; IPR000845; PNP UDP.
DR Pfam; PF01048; PNP_UDP_1_1.
DR TIGRFAWS; TIGR00107; deoD; 1.
DR TIGRFAWS; TIGR01718; Uridine-phosphatase; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
KW 3D-structure; Complete proteome; Direct protein sequencing;
KW Glycosyltransferase; Transferase.
FT INIT MET 0 0
FT STRAND 2 2
FT TURN 3 4
FT STRAND 5 5
FT TURN 9 10
FT STRAND 15 18
FT HELIX 22 32
FT STRAND 34 40
FT HELIX 42 44
FT STRAND 47 52
FT TURN 53 54
FT STRAND 55 60
FT HELIX 66 80
FT STRAND 84 93
FT TURN 96 97
FT TURN 100 101
FT STRAND 103 111
FT HELIX 115 119
FT TURN 120 122
FT STRAND 124 124
FT STRAND 127 128
FT HELIX 131 143
FT TURN 144 145
FT STRAND 148 155
FT TURN 164 164
FT HELIX 165 172

FT TURN 173 174
 FT STRAND 177 179
 FT HELIX 182 192
 FT TURN 193 193
 FT STRAND 195 205
 FT TURN 206 208
 FT STRAND 211 211
 FT HELIX 214 219
 FT TURN 223 236
 SQ SEQUENCE 238 AA; 25618 MW; C39ACFC1C817AC83E CRC64;

Query Match 99.3%; Score 1217; DB 1; Length 238;
 Best Local Similarity 99.6%; Pred. No. 4.9e-92;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ATPHINAEAGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISVM 61
 DB 1 ATPHINAEAGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISVM 60
 QY 62 GHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVLVIGMGACTDSKVNRI 121
 DB 61 GHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVLVIGMGACTDSKVNRI 120
 QY 122 KDHDFAAIAEDFDMVRNAVDAKAALGIDARVGNLFSADLFYSPDGEFVNNMEKYGIILGVE 181
 DB 121 KDHDFAAIAEDFDMVRNAVDAKAALGIDARVGNLFSADLFYSPDGEFVNNMEKYGIILGVE 180
 QY 182 EAAGIYGVAAEFAGKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239
 DB 181 EAAGIYGVAAEFAGKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 238

RESULT 5
 Q8ZJ02 PRELIMINARY; PRT; 239 AA.
 ID Q8ZJ02
 AC Q8ZJ02; Q7CAT5;
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DT 25-OCT-2004 (T-EMBLrel. 28, Last annotation update)
 DE Purine nucleoside phosphorylase (EC 2.4.2.1).
 GN Name-decd; Synonyms=pup; OrderedLocusNames=STV4921, t4613;
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18";
 RL Nature 413:848-852(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ty2 / ATCC 700931;
 RX MEDLINE=22531367; PubMed=12644504;
 RX DOI=10.1128/JB.185.7.2330-2337.2003;
 RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18";
 RL J. Bacteriol. 185:2330-2337(2003).
 DR EMBL; AL627284; CAD03405.1; -.
 DR EMBL; AE016849; AA072045.1; -.
 DR HSSP; P09743; 1PR6.

DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
 DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.
 DR GO; GO:0009116; P:nucleoside metabolism; IEA.
 DR Pfam; PF01048; PNP_UDP_1; 1.
 DR TIGRFAMS; TIGR00107; deod; 1.
 DR PROSITE; PS01232; PNP_UDP_1; 1.
 KW Complete proteome; Glycosyltransferase; Transferase.
 SQ SEQUENCE 239 AA; 25979 MW; 7A4AFD72C646859E CRC64;

Query Match 96.9%; Score 1187; DB 2; Length 239;
 Best Local Similarity 96.7%; Pred. No. 1.4e-89;
 Matches 231; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MATPHINAEAGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
 DB 1 MATPHINAEAGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
 QY 61 MHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVLPHVKLRDVLVIGMGACTDSKVNRI 120
 DB 61 MHGVGIPSCSIYTKELITDFGVKKIIRVSGGAVRMDVKLRDVLVIGMGACTDSKVNRI 120
 QY 121 FKDHDFAAIAEDFDMVRNAVDAKAALGIDARVGNLFSADLFYSPDGEFVNNMEKYGIILGVE 180
 DB 121 FKDHDFAAIAEDFDMVRNAVDAKAALGIDARVGNLFSADLFYSPDGEFVNNMEKYGIILGVE 180
 QY 181 MEAAGIYGVAAEFAGKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239
 DB 181 MEAAGIYGVAAEFAGKALITCTVSDHIRTHTQTTAAERQTTFNMDIKIALESVLLGDKE 239

RESULT 6
 Q8ZJ07 PRELIMINARY; PRT; 239 AA.
 ID Q8ZJ07
 AC Q8ZJ07;
 DT 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
 DE Purine nucleoside phosphorylase (EC 2.4.2.1).
 GN Name-decd; OrderedLocusNames=STW4570;
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2;
 RX MEDLINE=21534948; PubMed=11677609; DOI=10.1038/35101614;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2";
 RL Nature 413:852-856(2001).
 RL EMBL; AE008915; AAL23385.1; -.
 DR HSSP; P09743; 1PR6.
 DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
 DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.
 DR GO; GO:0009116; P:nucleoside metabolism; IEA.
 DR Pfam; PF01048; PNP_UDP_1; 1.
 DR TIGRFAMS; TIGR00107; deod; 1.
 DR PROSITE; PS01232; PNP_UDP_1; 1.
 KW Complete proteome; Glycosyltransferase; Transferase.
 SQ SEQUENCE 239 AA; 25978 MW; 90ED901A61814796 CRC64;

Query Match 96.5%; Score 1182; DB 2; Length 239;
 Best Local Similarity 96.2%; Pred. No. 3.7e-89;
 Matches 230; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MATPHINAEAGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
 DB 1 MATPHINAEAGDPADVVLMPGDPPLRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60

RX DOI=10.1128/JB.184.16.4601-4611.2002;
 RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.D., Matson J.S., Blattner F.R.,
 RA Perry R.D.;
 RT "Genome sequence of *Yersinia pestis* KIM.";
 RN J. Bacteriol. 184:4601-4611(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=91001 / Biovar Mediaevalis;
 RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,
 RA Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,
 RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,
 RA Yang R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE013977; AAM87286.1; -;
 DR EMBL; AE017141; AAS63890.1; -;
 DR HSSP; P09743; 1PR6.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
 DR GO; GO:0009116; P:nucleoside metabolism; IEA.
 DR InterPro; IPR004402; DeoD.
 DR Pfam; PF01048; PNP_UDP_1; 1.
 DR TIGRFAMs; TIGR00107; deoD; 1.
 DR PROSITE; PS01232; PNP_UDP_1; 1.
 SQ SEQUENCE 241 AA; 26120 MW; 797D09F97072CFA9 CRC64;

 Query Match 90.3%; Score 1106; DB 2; Length 241;
 Best Local Similarity 90.7%; Pred. No. 6.6e-83;
 Matches 215; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

 Qy 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMGLFTGTGKRSIV 60
 Db 3 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMGLFTGTGKRSIV 62

 Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHVKLRDVGWGMGACTDSKNRIR 120
 Db 63 MGHGMGIPSCSIYAKELITDFGVKKIIRVSGCAGVLPVHVKLRDVGWGMGACTDSKNRMR 122

 Qy 121 FKDHFAAIAADFDMVRNAVDAAKALGIDARVGNLFSADLPYSPDGMFDMVEKYGILGVE 180
 Db 123 FKDHVAAIAADFEMTRNAVDAAKAKGVNRVGNLFSADLPYTPDPQMFDMVEKYGILGVE 182

 Qy 181 MEAGIYGVAAEFGAKALITCTVSDHRTTAAEROTTENDMIKIALESVLLGD 237
 Db 183 MEAGICGVAAEFGAKALITCTVSDHRTTAAEROTTENDMIKIALESVLLGD 239

 RESULT 10
 Q6D989 PRELIMINARY; PRT; 239 AA.
 AC Q6D989
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Purine nucleoside phosphorylase (EC 2.4.2.1).
 GN Name=DeoB; Synonym=pnp; OrderedLocusNames=ECA0730;
 OS *Erwinia carotovora* (subsp. atroseptica) (Pectobacterium atrosepticum).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pectobacterium.
 OX NCBI_TaxID=29471;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SCRI 1043 / ATCC BAA-672;
 RX PubMed=15263089; DOI=10.1073/pnas.0402424101;
 RA Bell K.S., Sebahia M., Pritchard L., Holden M.T.G., Hyman L.J.,
 RA Holeva M.C., Thomson N.R., Bentley S.D., Churcher L.J.C., Mungall K.,
 RA Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,
 RA Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,
 RA Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,
 RA Salmond G.P.C., Birch P.R.J., Parkhill J., Toth I.K.;

RT "Genome sequence of the enterobacterial phytopathogen *Erwinia*
 RT *carotovora* subsp. atroseptica and characterization of virulence
 RT factors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 101:11105-11110(2004).
 DR EMBL; BX950851; CAG73644.1; -;
 DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
 DR GO; GO:0009116; P:nucleoside metabolism; IEA.
 DR InterPro; IPR004402; DeoD.
 DR InterPro; IPR000845; PNP_UDP.
 DR Pfam; PF01048; PNP_UDP_1; 1.
 DR TIGRFAMs; TIGR00107; deoD; 1.
 DR PROSITE; PS01232; PNP_UDP_1; 1.
 KW Complete proteome; Glycosyltransferase; Transferase.
 SQ SEQUENCE 239 AA; 26038 MW; AC910A030E86170A CRC64;

 Query Match 89.4%; Score 1095; DB 2; Length 239;
 Best Local Similarity 88.7%; Pred. No. 5.2e-82;
 Matches 212; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

 Qy 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMGLFTGTGKRSIV 60
 Db 1 MATPHINAEMGDFADVLMFGDPLRAKYIAETFLDAREVNNVGMGLFTGTGKRSIV 60

 Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHVKLRDVGWGMGACTDSKNRIR 120
 Db 61 MGHGMGIPSCSIYAKELITDFGVKKIIRVSGCAGVLPVHVKLRDVGWGMGACTDSKNRMR 120

 Qy 121 FKDHFAAIAADFDMVRNAVDAAKALGIDARVGNLFSADLPYSPDGMFDMVEKYGILGVE 180
 Db 121 FKDHVAAIAADFDMVRNAVDAAKARDVSVRVGNLFSADLPYTPDPQMFDMVEKYGILGVE 180

 Qy 181 MEAGIYGVAAEFGAKALITCTVSDHRTTAAEROTTENDMIKIALESVLLGDKE 239
 Db 181 MEAGIYGVAAEFGAKALITCTVSDHRTTAAEROTTENDMIKIALESVLLGDNE 239

 RESULT 11
 Q8KRT5 PRELIMINARY; PRT; 238 AA.
 ID Q8KRT5
 AC Q8KRT5
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Purine nucleoside phosphorylase.
 GN Name=DeoB;
 OS *Xenorhabdus nematophilus*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Xenorhabdus.
 OX NCBI_TaxID=628;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA He H., Forst S.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19061;
 RA Orchard S.S., Goodrich-Blair H.;
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF525420; AAM91930.1; -;
 DR EMBL; AY556400; AAT40584.1; -;
 DR HSSP; P09743; 1PKE.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
 DR GO; GO:0009116; P:nucleoside metabolism; IEA.
 DR InterPro; IPR004402; DeoD.
 DR InterPro; IPR000845; PNP_UDP.
 DR Pfam; PF01048; PNP_UDP_1; 1.
 DR TIGRFAMs; TIGR00107; deoD; 1.
 DR PROSITE; PS01232; PNP_UDP_1; 1.
 SQ SEQUENCE 238 AA; 26006 MW; 682D984DE51441CA CRC64;

 Query Match 88.7%; Score 1086; DB 2; Length 238;


```
Best Local Similarity 88.2%; Pred. No. 2.9e-81;
Matches 209; Conservative 15; Mismatches 13; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db 1 MATPHINAEMGDFADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGTYGKRISV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIIVGAGACTDSKVNRI 120
Db 61 MGHGVGIPSCSIYAKELITDFGVKKIIRVSGCGAVSDVKIRDVVIGAGACTDSKVNRI 120

Qy 121 FKDHDFAAIADFDVLRHAAVDAAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGVE 180
Db 121 FKDHDFAAIADFDVLRHAAVDAAKAKININRVGNIFSDVLFYSPDGMFVMEKYGILGVE 180

Qy 181 MEAAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 237
Db 181 MEAAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGD 237

RESULT 12
DEOD KLEPN STANDARD; PRT; 239 AA.
AC Q59482;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Purine nucleoside phosphorylase (EC 2.4.2.1) (Inosine phosphorylase)
DE (PNP)
GN Name=deod; Synonyms=pnp;
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LF1202;
RA Takehara M., Ling F., Izawa S., Inoue Y., Kimura A.;
RT "Molecular cloning and nucleotide sequence of purine nucleoside
RT phosphorylase and uridine phosphorylase genes from Klebsiella sp.";
RL Biosci. Biotechnol. Biochem. 59:1987-1990(1995).
CC -1- FUNCTION: Cleavage of guanosine or inosine to respective bases and
CC sugar-1-phosphate molecules.
CC -1- CATALYTIC ACTIVITY: Purine nucleoside + phosphate = purine +
CC alpha-D-ribose 1-phosphate.
CC -1- SUBUNIT: Homohexamer (By similarity).
CC -1- SIMILARITY: Belongs to the PNP/UDP phosphorylase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; X87881; CAA61136.1; --
CC HSP; P09743; 1K9S
CC InterPro; IPR004402; Deod.
CC Pfam; PF01048; PNP_UDP_1; 1.
CC TIGRFAMs; TIGR00107; deod; 1.
CC TRFAMs; TRF01718; Uridine-phosphatase; 1.
CC PROSITE; PS01232; PNP_UDP_1; 1.
CC Glycosyltransferase; Transferase.
SQ SEQUENCE 239 AA; 26196 MW; BC7514EDF5B27F5 CRC64;

Query Match 88.6%; Score 1085; DB 1; Length 239;
Best Local Similarity 90.4%; Pred. No. 3.5e-81;
Matches 217; Conservative 7; Mismatches 14; Indels 2; Gaps 2;

Qy 1 MATPHINAEMGDFADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKIS 59
```

```
Db 1 MATPHINAEMGDFADVLMPCGDLPRANSIAETFLDAREVNNVRGMLGFTGTYGRRKIS 60
Qy 60 VMGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIIVGAGACTDSKVNRI 119
Db 61 VMGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVREDVKLRVVIHGHGACTDSKVNRL 120

Qy 120 RFKDHDFAAIADFDVLRHAAVDAAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGV 179
Db 121 RFKDHDFPAIADFDVLRHAAVDAAKALGVARRSNIFSDLFYTPDPSPFVMEKYGILGV 179

Qy 180 EMEAAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGDKE 239
Db 180 EMEAAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLRDKE 239

RESULT 13
Q7N930 PRELIMINARY; PRT; 238 AA.
AC Q7N930;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Purine-nucleoside phosphorylase (Inosine phosphorylase) (PNP).
GN Name=deod; OrderedLocNames=plu0522;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-P.,
RA Dassa E., Derose R., Berzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Siguer P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens.";
RL Nat. Biotechnol. 21:1307-1313(2003).
RX EMBL; BX571860; CAEL2817.1; --
DR HSP; P09743; 1K91.
DR Photolista; plu0522; --
DR GO; GO:0003824; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deod.
DR InterPro; IPR000845; PNP_UDP.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
KW Complete proteome.
SQ SEQUENCE 238 AA; 26114 MW; 13651F7C82C49E9A CRC64;

Query Match 87.1%; Score 1067; DB 2; Length 238;
Best Local Similarity 85.7%; Pred. No. 1e-79;
Matches 204; Conservative 18; Mismatches 16; Indels 0; Gaps 0;

Qy 1 MATPHINAEMGDFADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGTYGKRKISV 60
Db 1 MATPHINAEMGDFADVLMPCGDLPRAKYIAETFLDAREVNNVRGMLGFTGTYGKRPLSV 60

Qy 61 MGHGVGIPSCSIYTKELITDFGVKKIIRVSGCGAVLPHVKLRDVIIVGAGACTDSKVNRI 120
Db 61 MGHGVGIPSCSIYAKELITDFGVKKIIRVSGCGSVRHDVKLRDVIIVGAGACTDSKVNRI 120

Qy 121 FKDHDFAAIADFDVLRHAAVDAAKALGIDARVGNLFSADLFYSPDGMFVMEKYGILGVE 180
Db 121 FKDHDFAAIADFDVLRHAAVDAAKAKININRVGNIFSVELFYTDPQLFDIMEKYGILGVE 180

Qy 181 MEAAGIYGVAAEFCAKALITCTVSDHRTHTQTTAAERQTTFNDMIKIALESVLLGDK 238
```

```
Db 181 MEAGFYGVAAYGAKALTICTVSDHIRTKEKLTAEERQTTFNEMIBIALESVLLGDK 238

RESULT 14
ID Q9CLE6 PRELIMINARY; PRT; 238 AA.
AC Q9CLE6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Deod.
GN Name=deod; OrderedLocusNames=PM1291;
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
OC NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100; DOI=10.1073/pnas.051634598;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR EMBL; AE006168; AAK03375.1; -.
DR HSSP; P09743; 1PR4.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004731; F:purine-nucleoside phosphorylase activity; IEA.
DR GO; GO:0009116; P:nucleoside metabolism; IEA.
DR InterPro; IPR004402; Deod.
DR InterPro; IPR000845; PNP_UDP.
DR Pfam; PF01048; PNP_UDP_1; 1.
DR TIGRFAMs; TIGR00107; deod; 1.
DR PROSITE; PS01232; PNP_UDP_1; 1.
KW Complete proteome.
SQ SEQUENCE 238 AA; 25951 MW; BFC27C2971833C05 CRC64;

Query Match 84.8%; Score 1039; DB 2; Length 238;
Best Local Similarity 86.0%; Pred. No. 2.1e-77;
Matches 202; Conservative 12; Mismatches 21; Indels 0; Gaps 0;

QY 3 TPHINAEWGDFADVLMPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISVMG 62
DB 2 TPHINAPAGAFADVLMPGDPLRAKYIAETFLQDVKEITNVRNMLGFTGYGKRKISVMG 61
QY 63 HGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHKLRDVVIGMGACTDSKVNIRFK 122
DB 62 HGMGIPSCSIYTKELITEYGVKKIIRVSGCAGVRNDVKVRDVIIGLACTDSKVNIRFK 121
QY 123 DHDFAAIADFDVNRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVEME 182
DB 122 NHDFAAIADFDMTWAAVQAQAKAGLVHVGNLFSADLFYTPDVMFDMVEKYGILGVEME 181
QY 183 AAGIYGVAABFGAKALTICTVSDHIRTKEKLTAEERQTTFNEMIBIALESVLLGD 237
DB 182 AAGIYGVAABFGAKALTICTVSDHIRTKEKLTAEERQTTPEERQTTFNEMIBIALESVLIGD 236

RESULT 15
Q65RA4 PRELIMINARY; PRT; 238 AA.
ID Q65RA4;
AC Q65RA4;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Deod protein.
GN Name=deod; ORFNames=MS1899;
OS Mannheimia succiniciproducens MBEL55E.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Mannheimia.
OC NCBI_TaxID=221988;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MBEL55E;
```

```
RA Hong S.H., Kim J.S., Lee S.Y., In Y.H., Choi S.S., Rih J.-K.,
RA Kim C.H., Jeong H., Hur C.-G., Kim J.G.;
RT "The genome sequence of the capnophilic rumen bacterium Mannheimia
RT succiniciproducens.";
RL Nat. Biotechnol. 0:0-0(2004).
DR EMBL; AE016827; AAU38506.1; -.
SQ SEQUENCE 238 AA; 25840 MW; 6A40A3B9C1E744BB CRC64;

Query Match 83.3%; Score 1021; DB 2; Length 238;
Best Local Similarity 84.3%; Pred. No. 6.3e-76;
Matches 199; Conservative 19; Mismatches 18; Indels 0; Gaps 0;

QY 3 TPHINAEWGDFADVLMPGDPLRAKYIAETFLDAREVNNVRGMLGFTGYGKRKISVMG 62
DB 2 TPHINAPAGAFADVLMPGDPLRAKYIAETFLNAKEVTNVRNMLGFTGYGKRPSVMG 61
QY 63 HGVGIPSCSIYTKELITDFGVKKIIRVSGCAGVLPVHKLRDVVIGMGACTDSKVNIRFK 122
DB 62 HGMGIPSCSIYTKELITEYGVKKIIRVSGCAGVRNDVKVRDVIIGLACTDSKVNIRFK 121
QY 123 DHDFAAIADFDVNRNAVDAAKALGIDARVGNLFSADLFYSPDGMFDMVEKYGILGVEME 182
DB 122 DNDFAAIADFDMTQAAVQAQAKAGIYRVGNLFSADLFYTPDVMFDMVEKYGILGVEME 181
QY 183 AAGIYGVAABFGAKALTICTVSDHIRTKEKLTAEERQTTFNEMIBIALESVLLGDK 238
DB 182 AAGIYGVAABFGAKALTICTVSDHIRTKEKLTAEERQTTFNEMIBIALESVLLGDQ 237

Search completed: July 18, 2005, 21:39:55
Job time : 177 secs
```